

SELF DESCRIPTION

QUESTIONNAIRE - I

SDQ I

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

Contents

Tables	vii
Figures	ix
Chapter 1. Introduction	1
Theoretical Basis of the SDQ	1
General Description	1
Overview of Empirical Research	3
Reliability and Factor Structure	3
Construct Validity	4
Changes in Self-Concept	4
Chapter 2. Administration and Scoring	9
Administration Instructions	9
Scoring Instructions	10
Individual Scale Raw Scores	11
Total Scores	11
Converting Raw Scores to Normative Scores	12
Control Scores	12
Chapter 3. Norms Development and Interpretation	19
Norms Development and Standardization	19
Nonnormalized Versus Normalized <i>T</i> Scores	19
Interpretation	20
Total Scores	23
General-Self Scale	23
Individual Scales	24
Control Scores	24
Discussion	26
Chapter 4. Multidimensional Assessment of Self-Concept	27
The Shavelson Model	27
Development and Factor Structure of the SDQ-I	29
Exploratory Factor Analyses	30
Confirmatory Factor Analyses	32
In Search of a General Self-Concept	39
Addition of a General-Self Scale to the SDQ-I	41
Summary and Implications	43
Chapter 5. Reliability	45
Internal Consistency Estimates	45
Stability and Systematic Changes	49
Chapter 6. Self-Concept and Academic Achievement	53
Self-Concept/Achievement Correlations	53
Sex Differences in Self-Concept/Achievement Relations	57
Chapter 7. Frame of Reference Effects on Self-Concept and Achievement	61
The Big Fish Little Pond Effect	61
Description of the Model	62
Support for the Model	62
Discussion	68
Internal/External Frame of Reference Model	69
Correlations Between Reading and Math Self-Concepts	71
The Achievement/Self-Concept Relationship	74
Inferred Self-Concepts	75
Discussion	76

Chapter 8. Self-Attributions for Academic Success and Failure	79
Correlations Between SDQ-I and Sydney Attribution Scale	80
The Self-Serving Effect.....	84
Effect of Outcome on Academic Attributions.....	85
Individual Differences.....	88
Discussion	92
Summary and Implications	94
Chapter 9. Inferred Self-Concepts	95
Theoretical Basis	95
Self-Other Agreement in SDQ Research	96
Summary and Implications	101
Chapter 10. Effects of Intervention	105
Methodological Issues in Intervention Studies.....	105
Study of the Outward Bound Standard Course.....	106
Study of the Outward Bound Bridging Course	106
Sample and Design	108
Intervention Effects.....	109
Discussion	114
Chapter 11. Age and Sex Effects	117
Tests of Factorial Invariance	117
Mean Self-Concept Responses	119
Research Review	119
SDQ-I Analyses.....	120
Summary and Implications	123
Chapter 12. Cross-National Comparisons	125
Tests of Factorial Invariance	127
Tests of Group and Sex Differences.....	127
Summary and Implications	128
Chapter 13. The Bias of Negative Items	133
Age Effects.....	134
Effects of Reading Ability.....	135
Summary and Implications	136
Chapter 14. Relationship Between Responses to the SDQ-I and Other Instruments	141
SDQ-III Research.....	142
Relationships with the <i>Coopersmith Self-Esteem Inventory</i>	143
Stability and Validity	143
Relationships with the Harter Perceived Competence Scale.....	146
Methods	147
Results and Discussion.....	147
Summary and Implications	151
Chapter 15. Summary of Theoretical and Empirical Research	153
Theoretical Basis	153
Construct Validation	154
References	155
Appendices.....	165

Tables

Table 1. Exploratory Factor Analysis for the Total Normative Sample.....	31
Table 2. Exploratory Factor Analysis by Grade Level	33
Table 3. Exploratory Factor Analysis Including the General-Self Scale.....	42
Table 4. Internal Consistency Coefficients for the Total Normative Sample by Scale.....	45
Table 5. Item Statistics by Scale	46
Table 6. Stability and Change in SDQ-I Scale Scores over a Six-Month Interval.....	51
Table 7. Correlations Between SDQ-I Scales and Academic Performance Measures.....	54
Table 8. Correlations Between Math and Reading/Verbal Self-Concepts in Studies Employing the SDQ-I, SDQ-II, and SDQ-III.....	73
Table 9. Path Coefficients for Testing the Internal/External Frame of Reference Model.....	75
Table 10. Path Coefficients for Testing the Internal/External Frame of Reference Model with Inferred Self-Concept Ratings.....	77
Table 11. Correlations Between Sydney Attribution Scale, SDQ-I Scales, Inferred Self-Concepts, and Academic Achievement Measures.....	82
Table 12. Effect of Level of Achievement and the Three Sydney Attribution Scale Facets on the Internality/Externality of Academic Self-Attributions.....	86
Table 13. Correlations Between the Self-Serving Effect and Self-Concepts, Inferred Self-Concepts, and Achievement Scores	89
Table 14. Correlations Between Multiple Facets of Self-Concept and Self-Concepts Inferred by Significant Others in Eight Studies.....	98
Table 15. Correlations Between Multiple Facets of Self-Concept and Self-Concepts Inferred by Significant Others for the SDQ-III.....	100
Table 16. Means and Standard Deviations for Self-Concept and Achievement Scores at Time 1, Time 2, and Time 3 for Each Year of the Study.....	110
Table 17. Repeated Measures ANOVAs Comparing Control Interval Changes with Experimental Interval Changes for Six Sets of Variables	113
Table 18. Goodness of Fit Indices for the Factor Models of Factorial Invariance Across SDQ-I Responses for Male and Female Fifth Graders.....	118
Table 19. Effects of Age and Sex on SDQ-I Factors Represented by Unweighted Scale Scores (UWS) and Factor Scores (FS).....	123

Table 20.	Goodness of Fit Indices for the Factor Models of Factorial Invariance Across Responses by Australian and English Preadolescents.	126
Table 21.	Model 12 Parameter Estimates for Tests of Sex, Group, and Sex x Group Interaction Effects for Australian and English School Children	130
Table 22.	Goodness of Fit Indices for the Confirmatory Factor Analysis Models Describing Tests of Sex, Group, and Sex x Group Interaction Effects for Australian and English School Children.	132
Table 23.	Investigation of Negative Item Bias: LISREL Maximum Likelihood Estimates for the Self-Concept Factors, Negative Item Factor, and Reading Ability Factor	138
Table 24.	Correlations among SDQ-I, the <i>Coopersmith Self-Esteem Inventory</i> , and Achievement Scores for Time 1 and Time 2	144
Table 25.	Summary of Factor Analyses of Responses to the SDQ-I and the Perceived Competence Scale (PCS)	148
Table 26.	Correlations Between SDQ-I and Perceived Competence Scale (PCS) and Academic Motivation Measures in Reading (RAch) and Mathematics (MAch).....	149
Appendix A-1.	Percentiles and <i>T</i> Scores for the Total Normative Sample	165
Appendix A-2.	Percentiles and <i>T</i> Scores: Males, Grades 2-4	166
Appendix A-3.	Percentiles and <i>T</i> Scores: Females, Grades 2-4	167
Appendix A-4.	Percentiles and <i>T</i> Scores: Males, Grades 5-6	168
Appendix A-5.	Percentiles and <i>T</i> Scores: Females, Grades 5-6	169
Appendix B.	Percentiles and <i>T</i> -Score Equivalents for the Experimental Control Scores	170

Figures

Figure 1.	SDQ-I Questionnaire.	2
Figure 2.	Physical Abilities Scale	5
Figure 3.	Physical Appearance Scale	5
Figure 4.	Peer Relations Scale	5
Figure 5.	Parent Relations Scale.....	6
Figure 6.	Reading Scale	6
Figure 7.	Mathematics Scale	6
Figure 8.	General-School Scale.....	7
Figure 9.	General-Self Scale	7
Figure 10a.	Scoring Example: Score Calculation and Summary from an SDQ-I Scoring and Profile Booklet	13
Figure 10b.	Scoring Example: <i>T</i> -Score Profile from an SDQ-I Scoring and Profile Booklet	14
Figure 10c.	Scoring Example: Control Score Calculation from an SDQ-I Scoring and Profile Booklet.....	17
Figure 11.	Higher Order Factor Structure of the SDQ-I	38
Figure 12.	Path Models of Relationships Among Gender, Verbal and Math Achievement Indicators, and Reading and Math Self-Concepts.....	59
Figure 13.	Theoretical Model of the External Frame of Reference Hypothesis	63
Figure 14.	Relationships Between SES and Academic Self-Concepts and Inferred Academic Self-Concepts.....	64
Figure 15.	Path Models of Relationships Among SES, Academic Ability, and Student Self-Concepts	66
Figure 16.	Path Model of Relationships Among Achievements and Reading and Math Self-Concepts: Internal/External Frame of Reference Model.....	71
Figure 17.	Outcome x Cause Interaction: High and Low Ability Students from the Total Group.....	87
Figure 18.	Grade and Sex Effects for Ten SDQ-I Scores in Grades 2 - 5	121

Chapter 1.

Introduction

Theoretical Basis of the SDQ

Self-concept has been used to explain a wide range of behaviors, and 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 previous literature on self-concept have noted important shortcomings, especially the lack of a strong theoretical basis and the poor quality of instrumentation. To remedy this situation, Shavelson, Hubner, and Stanton (1976) reviewed specific criteria for evaluating self-concept measures and proposed a multifaceted, hierarchical model (see Chapter 4). This model served as the basis for the *Self-Description Questionnaire-I* (SDQ-I) and its two companion instruments, the SDQ-II and the SDQ-III.

The SDQ was originally developed to measure self-concept in four nonacademic areas (Physical Ability, Physical Appearance, Peer Relations, and Parent Relations) and three academic areas (Reading, Mathematics, and General-School) and was subsequently revised to include a General-Self scale. The SDQ was also used to test several specific hypotheses from the Shavelson model and led to the revision of the model (Marsh, Byrne, & Shavelson, 1988; Marsh & Shavelson, 1985; Shavelson & Marsh, 1986). In this sense the SDQ-I, SDQ-II, and SDQ-III are all the result of a dynamic and ongoing interplay between theory and empirical research.

The identification of theoretically consistent and distinct facets of self-concept and their structure is, at least initially, a prerequisite to the study of how self-concept is related to other constructs. Consequently, early research with the SDQ-I focused on the internal characteristics of self-concept, particularly its facets and their organization. More recent SDQ-I research has focused on the relationships between specific self-concept facets and other constructs, such as academic achievement and self-concepts inferred by significant others, and on the effects of interventions designed to alter self-concept.

General Description

The 76-item SDQ-I (see Figure 1, page 2) assesses four areas of nonacademic self-concept, three areas of academic self-concept derived from the Shavelson model, and a General-Self scale derived from the Rosenberg (1965, 1979) self-esteem scale. These eight scales reflect a child's self-ratings in various areas of self-concept. (For a definition of each scale and a list of the items comprising it, see Figures 2 through 9, pages 5-7.)

In completing the SDQ-I children are asked to respond to simple declarative sentences (e.g., "I'm good at mathematics," "I make friends easily") with one of five responses: False, Mostly False, Sometimes False/Sometimes True, Mostly True, True. Each of the eight SDQ-I scales contains eight positively worded items. An additional 12 items are negatively worded in order to disrupt positive response biases; however, these are not included in the self-concept scores since research has

6293/4-1 R

NQ 155 - 283
1
SET

SELF-DESCRIPTION QUESTIONNAIRE - I



Your Name: _____ Circle one: Boy Girl

School: _____ Grade: _____ Age: _____

Teacher: _____ Date: _____

This is a chance to look at yourself. It is not a test. There are no right answers, and everyone will have different answers. Be sure that your answers show how you feel about yourself. PLEASE DO NOT TALK ABOUT YOUR ANSWERS WITH ANYONE ELSE. We will keep your answers private and not show them to anyone.

When you are ready to begin, please read each sentence and choose an answer. (You may read quietly to yourself as I read aloud.) There are five possible answers for each question: "True," "False," and three answers in between. There are five boxes next to each sentence, one for each of the answers. The answers are written at the top of the boxes. Choose your answer to a sentence and make a check mark in the box under the answer you choose. DO NOT say your answer out loud or talk about it with anyone else.

Before you start, there are three examples below. A student, Bob, has already answered two of these sentences to show you how to do it. In the third example you must choose your own answer and put in your own check mark.

EXAMPLES

		FALSE	MOSTLY FALSE	SOME-TIMES FALSE/ SOME-TIMES TRUE	MOSTLY TRUE	TRUE		
1.	I like to read comic books	1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	1
	Bob checked the box under the answer "True." This means that he really likes to read comic books. If Bob did not like to read comic books very much, he would have answered "FALSE" or "MOSTLY FALSE."							
2.	In general, I am neat and tidy	2	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2
	Bob answered "SOMETIMES FALSE, SOMETIMES TRUE," because he is not very neat, but he is not very messy either.							
3.	I like to watch TV.	3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	3
	For this sentence you have to choose the answer that is best for you. First you must decide if the sentence is "TRUE," or "FALSE," or somewhere in between. If you really like to watch TV. a lot, you would answer "TRUE" by making a check mark in the last box. If you hate watching TV., you would answer "FALSE" by making a check mark in the first box. If your answer is somewhere in between, then you would choose one of the other three boxes.							

If you want to change an answer you have marked, you should cross out the check mark and put a new check mark in another box on the same line.

For all the sentences be sure that your check mark is on the same line as the sentence you are answering. You should have one answer and only one answer for each sentence. Do not leave out any of the sentences. Once you have started, PLEASE DO NOT TALK. Turn over the page and begin.

		FALSE	MOSTLY FALSE	SOME-TIMES FALSE/ SOME-TIMES TRUE	MOSTLY TRUE	TRUE	
1. I am good looking	1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1
2. I'm good at all SCHOOL SUBJECTS	2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2
3. I can run fast	3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	3
4. I get good marks in READING	4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	4
5. My parents understand me	5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5
6. I hate MATHEMATICS	6	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	6
7. I have lots of friends	7	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	7
8. I like the way I look	8	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	8
9. I enjoy doing work in all SCHOOL SUBJECTS	9	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9
10. I like to run and play hard	10	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	10
11. I like READING	11	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	11
12. My parents are usually unhappy or disappointed with what I do	12	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	12
13. Work in mathematics is easy for me	13	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	13

		FALSE	MOSTLY FALSE	SOME-TIMES FALSE/ SOME-TIMES TRUE	MOSTLY TRUE	TRUE	
14. I make friends easily	14	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	14
15. I have a pleasant looking face	15	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	15
16. I get good marks in all SCHOOL SUBJECTS	16	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	16
17. I hate sports and games	17	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	17
18. I'm good at READING	18	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	18
19. I like my parents	19	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	19
20. I look forward to MATHEMATICS	20	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	20
21. Most kids have more friends than I do	21	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	21
22. I am a nice looking person	22	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	22
23. I hate all SCHOOL SUBJECTS	23	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	23
24. I enjoy sports and games	24	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	24
25. I am interested in READING	25	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	25
26. My parents like me	26	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	26

		FALSE	MOSTLY FALSE	SOME-TIMES FALSE/ SOME-TIMES TRUE	MOSTLY TRUE	TRUE	
27. I get good marks in MATHEMATICS	27	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	27
28. I get along with kids easily	28	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	28
29. I do lots of important things	29	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	29
30. I am ugly	30	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	30
31. I learn things quickly in all SCHOOL SUBJECTS	31	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	31
32. I have good muscles	32	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	32
33. I am dumb at reading	33	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	33
34. If I have children of my own, I want to bring them up like my parents raised me	34	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	34
35. I am interested in MATHEMATICS	35	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	35
36. I am easy to like	36	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	36
37. Overall, I am no good	37	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	37
38. Other kids think I am good looking	38	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	38
39. I am interested in all SCHOOL SUBJECTS	39	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	39

		FALSE	MOSTLY FALSE	SOME-TIMES FALSE/ SOME-TIMES TRUE	MOSTLY TRUE	TRUE	
40. I am good at sports	40	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	40
41. I enjoy doing work in READING	41	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	41
42. My parents and I spend a lot of time together	42	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	42
43. I learn things quickly in MATHEMATICS	43	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	43
44. Other kids want me to be their friend	44	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	44
45. In general, I like being the way I am	45	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	45
46. I have a good looking body	46	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	46
47. I am dumb in all SCHOOL SUBJECTS	47	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	47
48. I can run a long way without stopping	48	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	48
49. Work in READING is easy for me	49	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	49
50. My parents are easy to talk to	50	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	50
51. I like MATHEMATICS	51	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	51
52. I have more friends than most other kids	52	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	52

629 212-1

		FALSE	MOSTLY FALSE	SOME-TIMES FALSE/ SOME-TIMES TRUE	MOSTLY TRUE	TRUE	
53. Overall I have a lot to be proud of	53	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	53
54. I'm better looking than most of my friends	54	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	54
55. I look forward to all SCHOOL SUBJECTS	55	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	55
56. I am a good athlete	56	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	56
57. I look forward to READING	57	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	57
58. I get along well with my parents	58	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	58
59. I'm good at MATHEMATICS	59	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	59
60. I am popular with kids of my own age	60	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	60
61. I can't do anything right	61	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	61
62. I have nice features like nose, and eyes, and hair	62	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	62
63. Work in all SCHOOL SUBJECTS is easy for me	63	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	63
64. I'm good at throwing a ball	64	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	64

		FALSE	MOSTLY FALSE	SOME-TIMES FALSE/ SOME-TIMES TRUE	MOSTLY TRUE	TRUE	
65. I hate READING	65	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	65
66. My parents and I have a lot of fun together	66	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	66
67. I can do things as well as most other people	67	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	67
68. I enjoy doing work in MATHEMATICS	68	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	68
69. Most other kids like me	69	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	69
70. Other people think I am a good person	70	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	70
71. I like all SCHOOL SUBJECTS	71	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	71
72. A lot of things about me are good	72	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	72
73. I learn things quickly in READING	73	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	73
74. I'm as good as most other people	74	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	74
75. I am dumb at MATHEMATICS	75	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	75
76. When I do something, I do it well	76	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	76

SELF-DESCRIPTION QUESTIONNAIRE-I



SCORING AND PROFILE BOOKLET

HERBERT W. MARSH

I

NAME: _____

DATE: _____

SCHOOL: _____

SEX: M F AGE: _____ GRADE: _____

Score Calculation and Summary

INDIVIDUAL SCALE SCORES: For each scale, write the scores for the items listed in the blanks beside the item numbers. Sum the item scores within each scale and write the total raw score in the blank provided below the item scores.

Physical Abilities	Physical Appearance	Peer Relations	Parent Relations	Reading	Mathematics	General-School	General-Self
Item (Mean)*	Item (Mean)*	Item (Mean)*	Item (Mean)*	Item (Mean)*	Item (Mean)*	Item (Mean)*	Item (Mean)*
3 --- (3.84)	1 --- (3.53)	7 --- (4.46)	5 --- (4.38)	4 --- (3.78)	13 --- (3.52)	2 --- (3.35)	29 --- (3.74)
10 --- (4.14)	8 --- (3.64)	14 --- (4.01)	19 --- (4.80)	11 --- (3.96)	20 --- (3.23)	9 --- (3.56)	45 --- (4.35)
24 --- (4.66)	15 --- (3.39)	28 --- (4.10)	26 --- (4.79)	18 --- (3.95)	27 --- (3.75)	16 --- (3.42)	53 --- (4.22)
32 --- (3.71)	22 --- (3.43)	36 --- (3.73)	34 --- (4.30)	25 --- (3.96)	35 --- (3.64)	31 --- (3.77)	67 --- (4.22)
40 --- (4.28)	38 --- (3.23)	44 --- (3.98)	42 --- (4.10)	41 --- (3.87)	43 --- (3.76)	39 --- (3.79)	70 --- (4.06)
48 --- (3.82)	46 --- (3.42)	52 --- (3.35)	50 --- (4.27)	49 --- (3.94)	51 --- (3.62)	55 --- (3.64)	72 --- (4.05)
56 --- (3.89)	54 --- (3.15)	60 --- (3.98)	58 --- (4.53)	57 --- (3.82)	59 --- (3.76)	63 --- (3.40)	74 --- (4.16)
64 --- (4.36)	62 --- (3.70)	69 --- (4.02)	66 --- (4.34)	73 --- (4.04)	68 --- (3.51)	71 --- (3.63)	76 --- (4.19)
RAW SCALE TOTALS							

TOTAL NONACADEMIC: Copy the Raw Scale Totals for Physical Abilities, Physical Appearance, Peer Relations, and Parent Relations into the blanks provided below. Sum these scores and divide by 4 to get the Total Nonacademic raw score.

$$\frac{\text{(Physical Abilities)} + \text{(Physical Appearance)} + \text{(Peer Relations)} + \text{(Parent Relations)}}{4} = \text{TOTAL NONACADEMIC Raw Score}$$

TOTAL ACADEMIC: Copy the Raw Scale Totals for Reading, Mathematics, and General-School into the blanks provided below. Sum these scores and divide by 3 to get the Total Academic raw score.

$$\frac{\text{(Reading)} + \text{(Mathematics)} + \text{(General-School)}}{3} = \text{TOTAL ACADEMIC Raw Score}$$

TOTAL SELF: Copy the Total Nonacademic and Total Academic raw scores into the blanks provided below. Sum these scores and divide by 2 to get the Total Self raw score.

$$\frac{\text{(Total Nonacademic)} + \text{(Total Academic)}}{2} = \text{TOTAL SELF Raw Score}$$

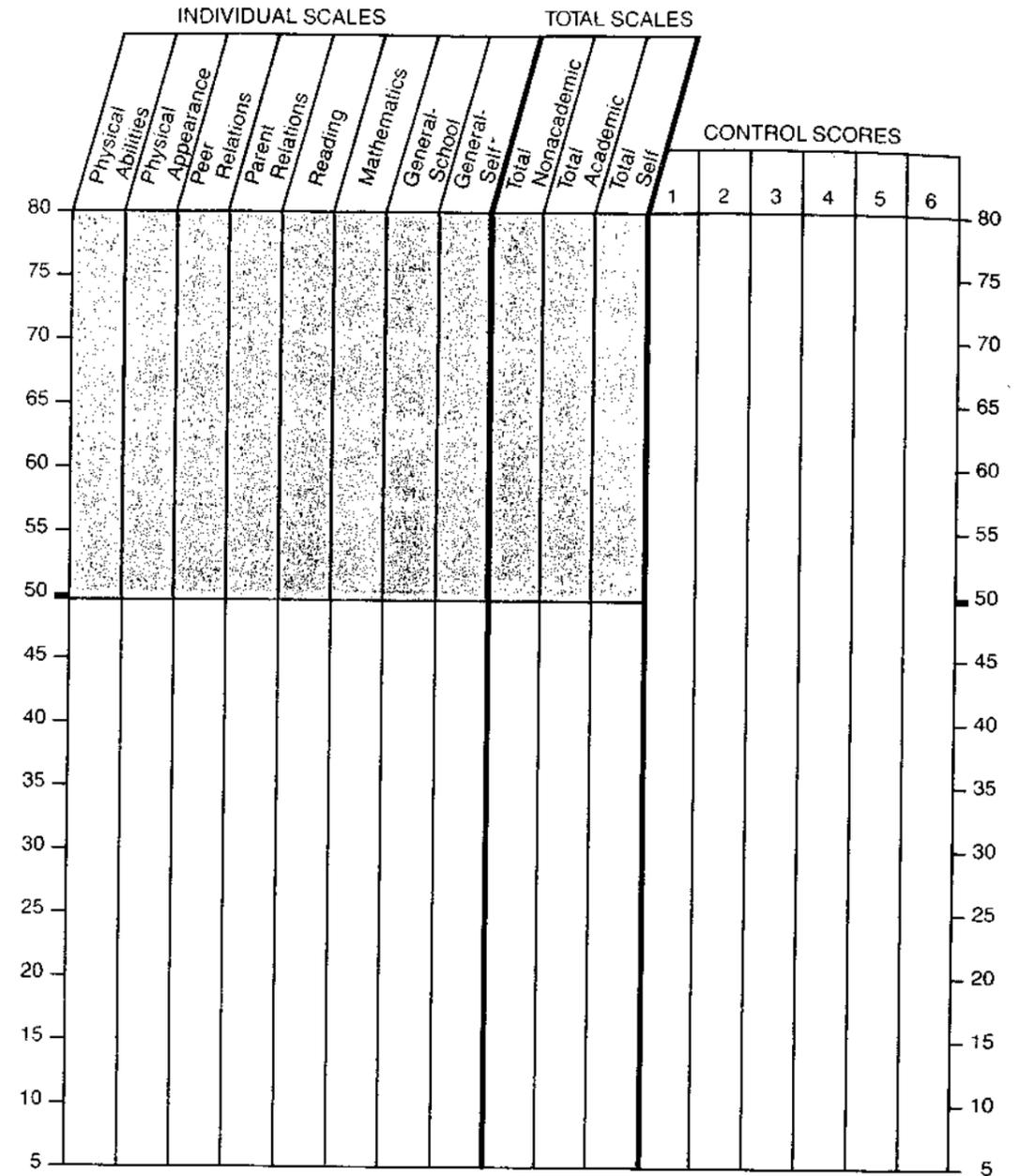
CONTROL SCORES (See Appendix A of the Manual for instructions on calculating Control raw scores.)

Control Score 1	Control Score 2	Control Score 3	Control Score 4	Control Score 5	Control Score 6
-----------------	-----------------	-----------------	-----------------	-----------------	-----------------

*Substitute the item mean for missing responses only if three or fewer responses are left blank.

T-Score Profile

actions: Transfer the raw scores for the individual and total scales (and control scores) 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 Appendices A and B of the Manual. Record these values in the spaces provided and plot the T scores on the profile.



Raw Scores: _____

Percentiles: _____

T Score: _____

*General-Self norms are not available for grades 2-4.

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.

Control Score Calculation

Control Score 1:
Consistency on
Correlated Item Pairs

Control Score 2:
Consistency on
Uncorrelated Item Pairs

Control Score 3 - Noncontingent Summary

Write the values of Control Score 2 and Control Score 1 in the appropriate blanks below. Subtract Control Score 1 from Control Score 2. Write the result in the blank labeled Control Score 3.

Item3	Item48	=	Item10	Item2	=
Item38	Item54	=	Item15	Item58	=
Item44	Item69	=	Item5	Item16	=
Item41	Item57	=	Item43	Item19	=
Item43	Item59	=	Item16	Item5	=
Item48	Item56	=	Item24	Item62	=
Item22	Item46	=	Item7	Item20	=
Item19	Item26	=	Item58	Item13	=
Item18	Item49	=	Item68	Item7	=
Item9	Item71	=	Item31	Item26	=
Item3	Item56	=	Item54	Item19	=
Item14	Item28	=	Item36	Item19	=
Item50	Item11	=	Item66	Item2	=
Item27	Item59	=	Item2	Item24	=
Item55	Item71	=	Item35	Item7	=
Item15	Item22	=	Item38	Item10	=
Item60	Item69	=	Item52	Item24	=
Item11	Item25	=	Item13	Item26	=
Item35	Item51	=	Item9	Item64	=
Item2	Item16	=	Item59	Item58	=
CONTROL SCORE 1:			CONTROL SCORE 2:		

Control Score 2	Control Score 1	CONTROL SCORE
-----------------	-----------------	---------------

Control Scores 4 and 5: Before entering the item values, reverse the direction of the scores so that 1 = True, 2 = Mostly True, 3 = Sometimes False/Sometimes True, 4 = Mostly False, and 5 = False.

Item17	x	8	=	Physical
Item21	x	8	=	Peers
Item33	x	8	=	Read
Item6	x	8	=	Math
Item23	x	8	=	School
Item30	x	8	=	Appearance
Item12	x	8	=	Parents
Item65	x	8	=	Read
Item75	x	8	=	Math
Item47	x	8	=	School
Total Absolute Value		+ 8		Total Signed Value
CONTROL SCORE 4		CONTROL SCORE 5		
Negativity Bias		Positivity Bias		

Control Score 6 - Individual Profile Variation:
Calculate the standard deviation of the original seven scales (Physical Abilities, Physical Appearance, Peer Relations, Parent Relations, Reading, Math, and General-School).

CONTROL SCORE 6

See Chapter 2 of the Manual for step-by-step directions for calculating Control Scores.

Chapter 1.

Introduction

Theoretical Basis of the SDQ

Self-concept has been used to explain a wide range of behaviors, and 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 previous literature on self-concept have noted important shortcomings, especially the lack of a strong theoretical basis and the poor quality of instrumentation. To remedy this situation, Shavelson, Hubner, and Stanton (1976) reviewed specific criteria for evaluating self-concept measures and proposed a multifaceted, hierarchical model (see Chapter 4). This model served as the basis for the *Self-Description Questionnaire-I* (SDQ-I) and its two companion instruments, the SDQ-II and the SDQ-III.

The SDQ was originally developed to measure self-concept in four nonacademic areas (Physical Ability, Physical Appearance, Peer Relations, and Parent Relations) and three academic areas (Reading, Mathematics, and General-School) and was subsequently revised to include a General-Self scale. The SDQ was also used to test several specific hypotheses from the Shavelson model and led to the revision of the model (Marsh, Byrne, & Shavelson, 1988; Marsh & Shavelson, 1985; Shavelson & Marsh, 1986). In this sense the SDQ-I, SDQ-II, and SDQ-III are all the result of a dynamic and ongoing interplay between theory and empirical research.

The identification of theoretically consistent and distinct facets of self-concept and their structure is, at least initially, a prerequisite to the study of how self-concept is related to other constructs. Consequently, early research with the SDQ-I focused on the internal characteristics of self-concept, particularly its facets and their organization. More recent SDQ-I research has focused on the relationships between specific self-concept facets and other constructs, such as academic achievement and self-concepts inferred by significant others, and on the effects of interventions designed to alter self-concept.

General Description

The 76-item SDQ-I (see Figure 1, page 2) assesses four areas of nonacademic self-concept, three areas of academic self-concept derived from the Shavelson model, and a General-Self scale derived from the Rosenberg (1965, 1979) self-esteem scale. These eight scales reflect a child's self-ratings in various areas of self-concept. (For a definition of each scale and a list of the items comprising it, see Figures 2 through 9, pages 5-7.)

In completing the SDQ-I children are asked to respond to simple declarative sentences (e.g., "I'm good at mathematics," "I make friends easily") with one of five responses: False, Mostly False, Sometimes False/Sometimes True, Mostly True, True. Each of the eight SDQ-I scales contains eight positively worded items. An additional 12 items are negatively worded in order to disrupt positive response biases; however, these are not included in the self-concept scores since research has

	FALSE	MOSTLY FALSE	SOME-TIMES FALSE/SOME-TIMES TRUE	MOSTLY TRUE	TRUE	
1. I am good looking	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	1
2. I'm good at all SCHOOL SUBJECTS	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2
3. I can run fast	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	3
4. I get good marks in READING	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	4
5. My parents understand me	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	5
6. I hate MATHEMATICS	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	6
7. I have lots of friends	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	7
8. I like the way I look	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	8
9. I enjoy doing work in all SCHOOL SUBJECTS	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9
10. I like to run and play hard	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10
11. I like READING	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	11
12. My parents are usually unhappy or disappointed with what I do	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	12
13. Work in mathematics is easy for me	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	13
	FALSE	MOSTLY FALSE	SOME-TIMES FALSE/SOME-TIMES TRUE	MOSTLY TRUE	TRUE	
14. I make friends easily	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	14
15. I have a pleasant looking face	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	15
16. I get good marks in all SCHOOL SUBJECTS	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	16
17. I hate sports and games	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	17
18. I'm good at READING	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	18
19. I like my parents	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	19
20. I look forward to MATHEMATICS	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	20
21. Most kids have more friends than I do	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	21
22. I am a nice looking person	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	22
23. I hate all SCHOOL SUBJECTS	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	23
24. I enjoy sports and games	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	24
25. I am interested in READING	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	25

Figure 1. SDQ-I Questionnaire

shown that young children and preadolescents do not give valid responses to these items (see Chapter 13).

The SDQ-I can be administered individually or in groups. No special administration training is required. Actual presentation of the test items requires only about 8 to 10 minutes; however, it usually takes an additional 5 to 10 minutes to read the instructions and answer any questions. Thus, the total testing time may vary from about 15 to 20 minutes, depending on the age of the children and the number of children being tested. (See Chapter 2 for complete administration instructions.)

The SDQ-I was originally intended for use in grades 4 through 6 (ages 8 through 12). However, results described in Chapter 11 suggest that the test is suitable for children as young as grade 2 and, with appropriate modifications, for older students up through high school and perhaps even college age. The SDQ-II and SDQ-III were developed for use with early and late adolescents, respectively; however, this Manual focuses primarily on SDQ-I.

Norms based on the responses of 3,562 students (New South Wales, Australia) in grades 2 through 6 are presented in Appendix A-1 for each of the SDQ-I scales and for total scores based on all academic items, all nonacademic items, and all items. Because responses vary systematically with grade level and sex, separate norms tables are also presented by sex and by grade range for grades 2 to 4 (Appendices A-2 and A-3) and grades 5 to 6 (Appendices A-4 and A-5), respectively. Several internal checks for random and biased responding have also been developed (see Chapter 3). Scoring instructions for these experimental control scores are presented in Chapter 2. Normative data for these scores are also available (see Appendix B).

Overview of Empirical Research

The SDQ-I is one in a series of three instruments designed to measure self-concepts for children and preadolescents (SDQ-I), young adolescents (SDQ-II), and late adolescents and young adults (SDQ-III). Although this Manual specifically describes the SDQ-I, relevant research conducted with older subjects responding to the SDQ-II or SDQ-III is also reviewed, particularly in subsequent chapters that focus on theoretical issues.

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 SDQ-I studies. The main empirical tools for this type of research are reliability and factor analytic studies and multitrait-multimethod analyses. The internal consistency reliability estimates for the various scales and total scores are all in the .80s and .90s, while the average correlation among the individual self-concept scales is relatively low (mean $r = .17$; see Chapter 5). Because self-concept in the Shavelson model is considered to be hierarchically ordered as well as multifaceted, recent advances in the application of hierarchical factor analysis were used to test the hierarchical structure of responses to the SDQ-I (see Chapter 4). Factor analysis was also used to test the invariance of the SDQ-I factor structure across diverse groups (see Chapters 4, 11, and 12). Though most research with the SDQ-I has been based on Australian samples, a cross-national study shows the generalizability of the SDQ-I factor structure, sex differences in specific SDQ-I scales, and mean responses to SDQ-I scales for a large sample of English school children (see Chapter 12). Numerous applications of multitrait-multimethod analyses have further demonstrated

the distinctiveness of the different factors (see Chapters 5, 9, and 10). These studies demonstrate that the SDQ-I reliably measures distinct facets of self-concept.

Construct Validity

Self-concept is a theoretical construct. It is therefore appropriate to use a construct validation approach to test the validity of the responses to the SDQ-I. Such an approach requires that SDQ-I responses be related to a 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-I were found to be related to sex, age, socioeconomic status, academic achievement, teacher ratings of achievement and inferred self-concept, peer ratings of inferred self-concept, student self-attributions for the perceived causes of their academic successes and failures, responses to other self-concept instruments, and experimental interventions designed to enhance self-concept. SDQ-I responses are systematically related to these external criteria in a way that is consistent with the theory, thus supporting the construct validity of the instrument.

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 possible 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, self-concept *is* stable over time. Despite claims to the contrary, very few programs to enhance self-concept have any measurable effect. There are several reasons for these failures. Some interventions fail because they are inherently weak, others because they are applied to such a small number of children that changes cannot be reliably assessed, and still others because self-concept instruments used do not measure the specific components of self-concept that are logically related to the intended effects of the intervention. An important area of SDQ-I research discussed in this Manual is the study of the effects of interventions on responses to the SDQ-I and the examination of alternative explanations for these effects (see Chapter 10).

The Physical Abilities scale measures a child's self-concept regarding his or her abilities in physical activities, sports, and games. Items in the Physical Abilities scale are:

Item No.	Item
3.	I can run fast.
10.	I like to run and play hard.
24.	I enjoy sports and games.
32.	I have good muscles.
40.	I am good at sports.
48.	I can run a long way without stopping.
56.	I am a good athlete.
64.	I am good at throwing a ball.

Figure 2. Physical Abilities Scale

Items of the Physical Appearance scale reflect a child's self-concept regarding his or her physical attractiveness as compared with others, and the perception of how others think he or she looks. Items in the Physical Appearance scale are:

Item No.	Item
1.	I am good looking.
8.	I like the way I look.
15.	I have a pleasant looking face.
22.	I am a nice looking person.
38.	Other kids think I am good looking.
46.	I have a good looking body.
54.	I am better looking than most of my friends.
62.	I have nice features like nose, and eyes, and hair.

Figure 3. Physical Appearance Scale

The Peer Relations scale measures the child's self-concept regarding his or her popularity with peers, how easily the child makes friends, and whether others want him or her as a friend. Items of the Peer Relations scale are:

Item No.	Item
7.	I have lots of friends.
14.	I make friends easily.
28.	I get along with kids easily.
36.	I am easy to like.
44.	Other kids want me to be their friend.
52.	I have more friends than most other kids.
60.	I am popular with kids my own age.
69.	Most other kids like me.

Figure 4. Peer Relations Scale

The Parent Relations scale reflects how well the child thinks he or she gets along with his or her parents, how well the child likes his or her parents, and the extent to which the child experiences parental acceptance and approval. Items of the Parent Relations scale are:

<u>Item No.</u>	<u>Item</u>
5.	My parents understand me.
19.	I like my parents.
26.	My parents like me.
34.	If I have children of my own, I want to bring them up like my parents raised me.
42.	My parents and I spend a lot of time together.
50.	My parents are easy to talk to.
58.	I get along well with my parents.
66.	My parents and I have a lot of fun together.

Figure 5. Parent Relations Scale

The Reading scale reflects the child's self-concept regarding his or her ability, enjoyment, and interest in reading. Items of the Reading scale are:

<u>Item No.</u>	<u>Item</u>
4.	I get good marks in READING.
11.	I like READING.
18.	I am good at READING.
25.	I am interested in READING.
41.	I enjoy doing work in READING.
49.	Work in READING is easy for me.
57.	I look forward to READING.
73.	I learn things quickly in READING.

Figure 6. Reading Scale

The Mathematics scale measures the child's self-concept regarding his or her ability, enjoyment, and interest in mathematics. Items of the Mathematics scale are:

<u>Item No.</u>	<u>Item</u>
13.	Work in MATHEMATICS is easy for me.
20.	I look forward to MATHEMATICS.
27.	I get good marks in MATHEMATICS.
35.	I am interested in MATHEMATICS.
43.	I learn things quickly in MATHEMATICS.
51.	I like MATHEMATICS.
59.	I am good at MATHEMATICS.
68.	I enjoy doing work in MATHEMATICS.

Figure 7. Mathematics Scale

The General-School scale measures the child's self-concept regarding his or her ability, enjoyment, and interest in school subjects. Items of the General-School scale are:

<u>Item No.</u>	<u>Item</u>
2.	I am good at all SCHOOL SUBJECTS.
9.	I enjoy doing work in all SCHOOL SUBJECTS.
16.	I get good marks in all SCHOOL SUBJECTS.
31.	I learn things quickly in all SCHOOL SUBJECTS.
39.	I am interested in all SCHOOL SUBJECTS.
55.	I look forward to all SCHOOL SUBJECTS.
63.	Work in all SCHOOL SUBJECTS is easy for me.
71.	I like all SCHOOL SUBJECTS.

Figure 8. General-School Scale

The General-Self scale reflects the child's perception of himself or herself as an affective, capable individual, proud of and satisfied with the way he or she is. Items of the General-Self scale are:

<u>Item No.</u>	<u>Item</u>
29.	I do lots of important things.
45.	In general, I like being the way I am.
53.	Overall I have a lot to be proud of.
67.	I can do things as well as most other people.
70.	Other people think I am a good person.
72.	A lot of things about me are good.
74.	I'm as good as most other people.
76.	When I do something, I do it well.

Figure 9. General-Self Scale

Chapter 2.

Administration and Scoring

Administration Instructions

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

- 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 so that generalizations described in this Manual may not be appropriate.
- Give a copy of the SDQ-I Questionnaire and a pencil with an eraser to each student. Help the students 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.*
- 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 puzzled at the end of the second paragraph, but this puzzlement 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 five boxes 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 arriving at an answer to Example 3, and most understand how to mark their answer. However, questions will not be allowed once administration of the test begins, so answer all questions now, and make sure the students understand how to respond.
- After all students have responded to Example 3, be sure that they *do not* turn the page until after you have read the next paragraph aloud. After you read the sentence "Do not leave out any of the sentences," add the following statement:

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.
- 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.**
- At this time if there are any questions, go to the individual student. If a student has trouble understanding a few words or expressions, paraphrase the expression as best you can without changing the meaning of the sentence. Ask the student to answer it as best 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.

Although problems in administering the SDQ-I are rare, several potential problems and solutions are presented below to assist the user.

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.

It is also possible that a student may mistakenly mark the answer to one or more items in the wrong place on the Questionnaire. The layout of the SDQ-I makes this unlikely, but if this happens, simply tell the student to crossout the incorrect response and substitute the correct one. If this has occurred for a large number of responses, it may be necessary to transfer the correct responses to a new Questionnaire.

Finally, 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-I 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.

Scoring Instructions

Responses to the SDQ-I may be scored conveniently using the SDQ-I Scoring and Profile Booklet. The Booklet provides for the calculation of individual scale raw scores, total raw scores (Nonacademic, Academic, and Total Self), and optional control scores. Calculation of each of the scores is described below.

Individual Scale Raw Scores

Complete the identifying information on the front cover of the SDQ-I Scoring and Profile Booklet by copying the information from the front of the child's SDQ-I Questionnaire.

Calculation of the individual raw scores 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 there is a column of item numbers that comprise the scale, blanks in which the child's item scores should be recorded, and item means (in parentheses). Using the child's Questionnaire, find the child's responses for items 3, 10, 24, and so on down the column for the Physical Abilities Scale. Convert the child's response to each item into one of the following scores: False = 1, Mostly False = 2, Sometimes False/Sometimes True = 3, Mostly True = 4, and True = 5. Write the appropriate score for each item in the blank after the item number. Repeat this procedure for the seven remaining individual scales and record the appropriate score for each item in the blanks provided after the item numbers.

Then simply sum the scores in each column to arrive at the individual scale raw scores. For example, sum the scores for Items 3, 10, 24, 32, 40, 48, 56, and 64 to arrive at the raw score for the Physical Abilities Scale. Write the sum for each scale in the blanks labeled "Individual Raw Scores" that appear below each column of item numbers.

If on the Questionnaire the child omits *three or fewer* responses, the mean response for the missing item should be substituted for the missing item score. Item means are listed in parentheses following the blanks next to their respective item numbers on the Score Calculation and Summary page. They are also listed in Table 5 (pages 46-48). (If there are four or more responses missing, the responses either should not be scored at all or 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 next whole number.

Users are cautioned to 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. In addition, as each individual scale score is summed, it should be checked by recalculating it to avoid errors in addition. Also, note that for each scale the lowest possible raw score is 8, and the highest possible raw score is 40.

Total Scores

The individual scale raw scores are used to calculate the Total Academic, Total Nonacademic, and Total Self raw scores.

Four of the individual scale raw scores (Physical Abilities, Physical Appearance, Peer Relations, and Parent Relations) are used to calculate the Total Nonacademic score. Three of the individual scale raw scores (Reading, Math, and General-School) are used to calculate the Total Academic score. These two composite scores are used in turn to compute the Total Self score.

To calculate the Total Nonacademic score, copy the raw scores for the Physical Abilities, Physical Appearance, Peer Relations, and Parent Relations scales in the appropriate blanks in the section labeled "Total Nonacademic" on the Score Calculation and Summary page. Then sum the four raw scores and write the total in the space labeled "Total." Finally, divide the total by four and write the result in

the space labeled "Total Nonacademic Raw Score." If the result is not a whole number, round it to the nearest whole number.

To calculate the Total Academic score, copy the raw scores for the Reading, Math, and General-School scales into the appropriate blanks in the section labeled "Total Academic." Then sum the three raw scores and write the total in the space labeled "Total." Finally, divide the total by three and write the result in the space labeled "Total Academic Raw Score." If the result is not a whole number, round it to the nearest whole number.

Finally, calculate the Total Self score. To do so, copy the Total Nonacademic and the Total Academic raw scores into the appropriate blanks in the section labeled "Total Self." Then sum the two composite scores and write the total in the space labeled "Total." Divide the total by two and write the result in the space labeled "Total Self Raw Score." If the result is not a whole number, round it to the nearest whole number.

Again, users are cautioned to double-check these calculations to avoid errors in addition or division. The possible range of scores can also serve as a check. For each total score (Total Nonacademic, Total Academic, and Total Self), the lowest possible raw score is 8, and the highest possible raw score is 40.

An example of a completely scored Score Calculation and Summary page from an SDQ-I Scoring and Profile Booklet is provided in Figure 10a.

Converting Raw Scores to Normative Scores

Raw scores for the SDQ-I individual scale and total raw scores may be converted to mid-interval percentile ranks and standard scores — in this case non-normalized *T* scores. *T* scores have a mean of 50 and a standard deviation of 10.

Normative comparisons are reported separately for males and females in grades 2 through 4 (Appendices A-2 and A-3) and 5 through 6 (Appendices A-4 and A-5), respectively. A combined norms table, based on responses by both males and females from grades 2 through 6, is provided (Appendix A-1) and may be useful in making group comparisons.

To use the norms tables, locate the appropriate row in the table for each raw score. Then find the column entries for the percentile or *T* score for that scale (e.g., Peer Relations percentiles in column 6). Enter these normative values in the spaces provided at the bottom of the *T*-Score Profile page of the Scoring and Profile booklet. Figure 10b, page 14, is an example of a completed *T*-Score Profile.

Control Scores

The SDQ-I provides checks for inappropriate responses in the form of experimental control scores. Although these control scores are optional, they should be calculated when there is a possibility that the child has not responded appropriately to the SDQ-I items.

In scoring the SDQ-I check for problems that might invalidate the results. Again, if more than four responses have been omitted from the Questionnaire, the responses either should not be scored at all or should be interpreted cautiously. If there are fewer than four missing responses, enter the mean response for the missing item or items. Also, if you have noted any unusual problems for a particular child (such

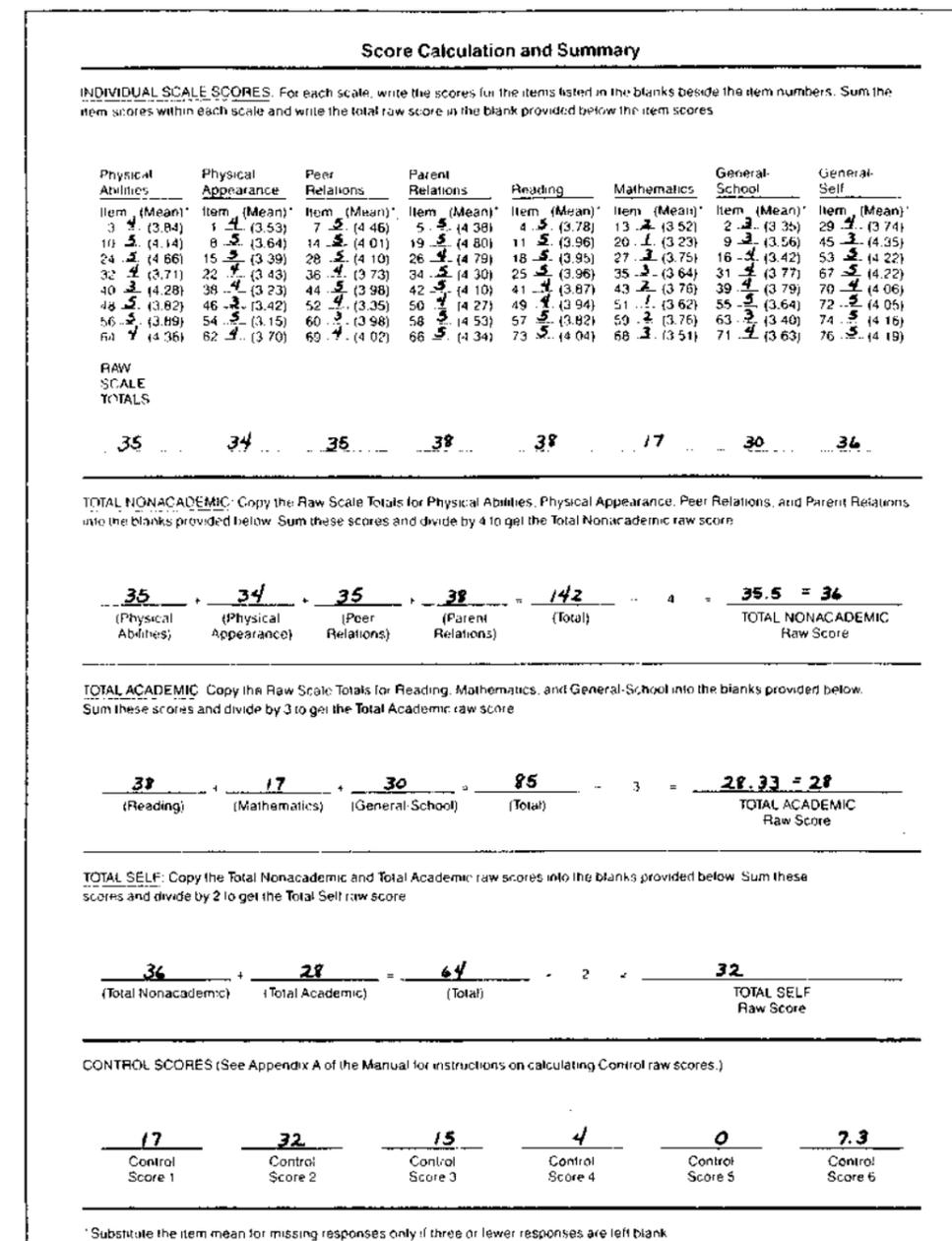


Figure 10a. Scoring Example: Score Calculation and Summary from an SDQ-I Scoring and Profile Booklet

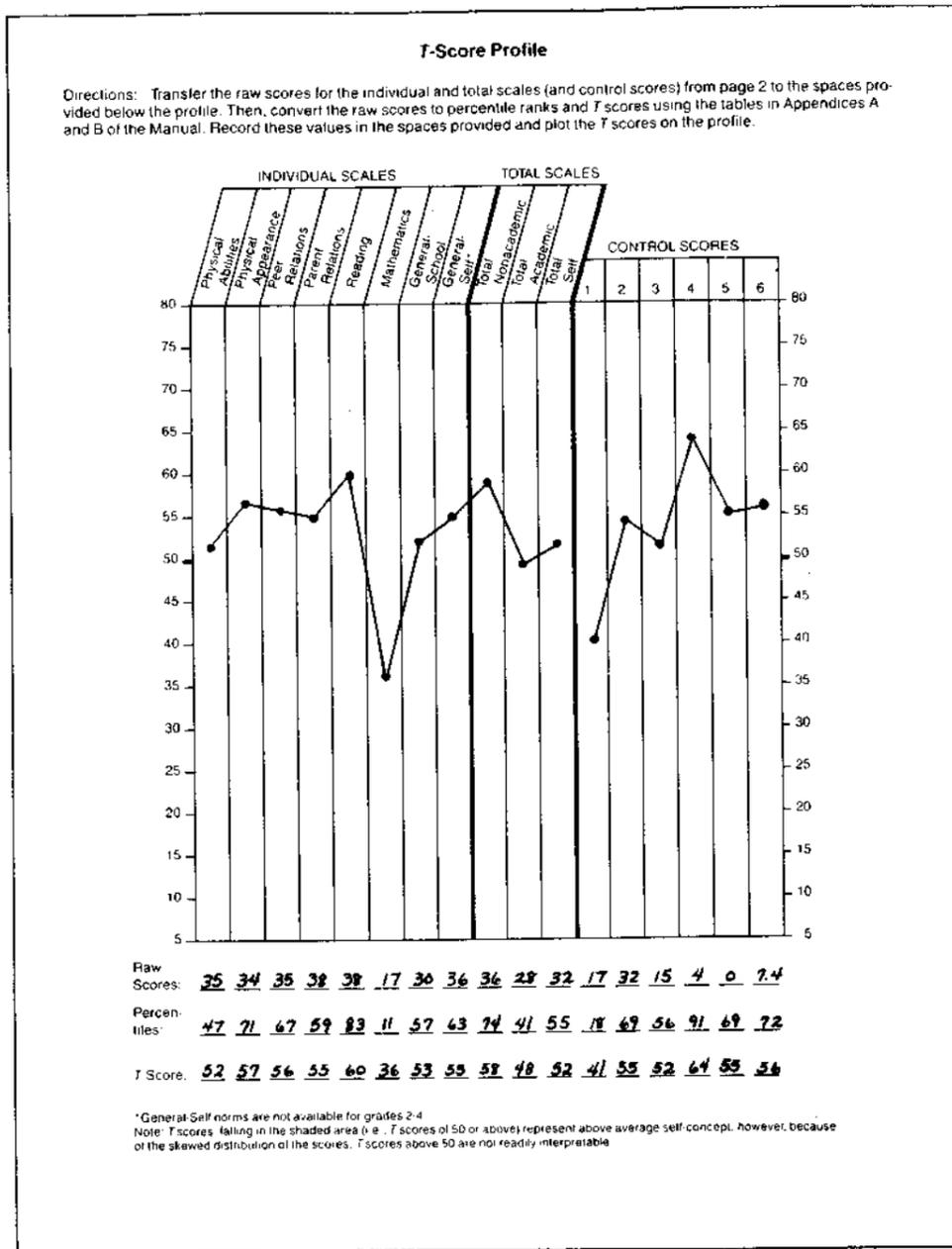


Figure 10b. Scoring Example: T-Score Profile from an SDQ-I Scoring and Profile Booklet

problems should be written on the front of the Questionnaire), the results either should not be scored at all or should be interpreted cautiously. A Control Score Calculation worksheet is provided on the last page of the Scoring and Profile Booklet. To use this worksheet to calculate the control scores, follow the procedures below:

Control Score 1: Inconsistency on Correlated Item Pairs. Control Score 1 is calculated by subtracting one item from another in item pairs and then summing the absolute (unsigned) results. The item pairs are listed in columns down the left side of the Control Score Calculation page of the Scoring and Profile Booklet.

First, find the child's response to each item and convert it into one of the following scores: False = 1, Mostly False = 2, Sometimes False/Sometimes True = 3, Mostly True = 4, and True = 5. Write the appropriate score for each item in the blank above the item number. Then find the item pair differences by subtracting the score for the second item in each pair from the score for the first item in the pair. Write the resulting difference scores for each item pair in the blank beside the item pair omitting any signs (negative or positive). Finally, sum the absolute difference scores by adding down the column and write the total in the space labeled "Control Score 1."

Control Score 2: Consistency on Uncorrelated Item Pairs. Control Score 2 is also calculated by subtracting one item score from another in pairs of items and then summing the absolute (unsigned) values of the differences. Follow the procedures outlined for Control Score 1, but record the item values and difference scores in the second column of blanks on the Control Score Calculation page. Write the sum of the difference scores in the blank labeled "Control Score 2."

Control Score 3: Noncontingent Summary. The Noncontingent Summary is the difference between Control Score 1 and Control Score 2. Write the values of these control scores in the appropriate blanks in the upper right corner of the Control Score Calculation page. Subtract Control Score 1 from Control Score 2 and write the result in the space labeled "Control Score 3."

Control Score 4: Negativity Bias. Control Scores 4 and 5 result from the same set of calculations, except that absolute (unsigned) values are summed for Control Score 4 and signed values are summed for Control Score 5. Begin calculating these control scores by finding the child's responses to the items listed in the Control Scores 4 and 5 section of the Control Score Calculation page of the Scoring and Profile Booklet. The items used to calculate Control Scores 4 and 5 are scored in the opposite direction from those used in earlier calculations (i.e., a high item score indicates low self-concept on that item); therefore, they must be converted to scores where True = 1, Mostly True = 2, Sometimes False/Sometimes True = 3, Mostly False = 4, and False = 5. Write each item score in the blank labeled with the appropriate item number. Then multiply each item score by 8 and write the result in the next column of blanks. Next, refer to the Score Calculation and Summary page of the Scoring and Profile Booklet to find the raw scores for the individual scales. Write the scale raw scores in the blanks labeled with the names of the appropriate scales. Note that the Reading and Math scale raw scores are used twice. Next, subtract each scale raw score from the value in the preceding blank (i.e., the item score multiplied by 8). Some of the resulting values will be negative. Write the result of each subtraction, including the sign (negative or positive), in the final column of blanks.

Finally, sum the absolute, unsigned values in the last column and write the sum in the blank labeled "Control Score 4."

Control Score 5: Positivity Bias. Sum the *signed values* in the last column of calculations used for Control Score 4. Write the result in the blank labeled "Control Score 5."

Control Score 6: Individual Profile Variation. Control Score 6 is the standard deviation of the child's individual scale raw scores, not including the General-Self scale. Control Score 6 can be calculated using a hand calculator which has a function key for computing standard deviations (using the n-1 option). To hand calculate Control Score 6, first sum the child's scale raw scores (from the Score Calculation and Summary page) for the Physical Abilities, Physical Appearance, Peer Relations, Parent Relations, Reading, Math, and General-School scales. Divide the sum of the scale raw scores by 7 to find the child's scale mean. Next subtract each scale raw score from the scale mean. The resulting values are the deviation scores. Square each deviation score (i.e., multiply each deviation score by itself). Sum the squared deviation scores and divide the sum by 6 (i.e., the number of scale scores minus 1). Finally, find the square root of the result of the preceding step. This square root is Control Score 6.

The control scores can then be transferred to the bottom of the Score Calculation and Summary page. Tables for converting control scores to mid-interval percentiles and normalized *T* scores are presented in Appendix B. Figure 10c is an example of a completed Control Score Calculation worksheet.

Control Score Calculation											
Control Score 1 Inconsistency on Correlated Item Pairs			Control Score 2 Consistency on Uncorrelated Item Pairs			Control Score 3 - Noncontingent Summary Write the values of Control Score 2 and Control Score 1 in the appropriate blanks below. Subtract Control Score 1 from Control Score 2. Write the result in the blank labeled Control Score 3.					
Item 3	Item 48	1	Item 10	Item 2	2	Control Score 2	Control Score 1	CONTROL SCORE 3			
4	5	1	5	3	2	32	17	15			
Item 36	Item 54	1	Item 15	Item 58	0	Control Scores 4 and 5: Before entering the item values, reverse the direction of the scores so the 1 = True, 2 = Mostly True, 3 = Sometimes True/Sometimes False, 4 = Mostly False, and 5 = False.					
4	5	1	5	5	0						
Item 44	Item 65	1	Item 5	Item 46	1						
5	4	1	5	4	1						
Item 41	Item 57	1	Item 43	Item 19	3	Item 17	Physical	40	35	5	
2	2	0	Item 16	Item 5	1	Item 21	Peers	32	35	-3	
Item 1	Item 59	0	Item 24	Item 62	1	Item 33	Read	40	38	2	
5	5	0	Item 17	Item 20	4	Item 6	Math	16	17	-1	
4	3	1	Item 2	Item 13	3	Item 23	School	24	30	-6	
Item 25	Item 26	1	Item 68	Item 7	2	Item 69	Appearance	40	34	6	
5	4	1	Item 31	Item 26	0	Item 12	Parents	32	38	-6	
Item 18	Item 49	1	Item 34	Item 19	0	Item 65	Read	40	38	2	
3	4	1	Item 36	Item 19	1	Item 75	Math	16	17	-1	
Item 9	Item 71	1	Item 6	Item 2	2	Item 47	School	32	30	2	
4	5	1	Item 35	Item 7	2	Total Absolute Value	34	Total Signed Value	0		
Item 3	Item 56	1	Item 38	Item 10	1	CONTROL SCORE 4	4.25 = 4	CONTROL SCORE 5	0		
5	5	0	Item 52	Item 24	1	Negativity Bias		Positivity Bias			
Item 14	Item 28	0	Item 13	Item 26	2	Control Score 6: Individual Profile Variation: Calculate the standard deviation of the original seven scales (Physical Abilities, Physical Appearance, Peer Relations, Parent Relations, Reading, Math, and General-School).					
4	5	1	Item 9	Item 64	1	7.32 = 7.4 CONTROL SCORE 6					
Item 50	Item 11	1	Item 59	Item 58	3	See Chapter 2 of the Manual for step-by-step directions for calculating Control Scores.					
3	2	1	Item 27	Item 59	1						
Item 27	Item 59	1	Item 2	Item 24	2						
5	4	1	Item 15	Item 22	1						
Item 55	Item 71	1	Item 80	Item 80	1						
5	4	1	Item 11	Item 25	2						
Item 15	Item 22	1	Item 2	Item 5	1						
3	4	1	Item 5	Item 5	0						
Item 80	Item 80	1	Item 11	Item 25	2						
5	5	0	Item 2	Item 5	1						
Item 11	Item 25	2	Item 2	Item 5	1						
3	1	2	Item 2	Item 5	3						
Item 35	Item 51	2	Item 2	Item 5	3						
3	4	1	CONTROL SCORE 1	CONTROL SCORE 2	32						
Item 2	Item 15	1									

Figure 10c. Scoring Example: Control Score Calculation from an SDQ-I Scoring and Profile Booklet

Chapter 3.

Norms Development and Interpretation

This chapter describes the standardization sample and the development of the norms contained in Appendices A and B and presents guidelines for interpreting the SDQ-I.

Norms Development and Standardization

The norms in this Manual are based on responses by 3,562 children from primary schools in New South Wales, Australia, collected between 1981 and 1983. Data were collected by the author and one of his colleagues in a series of research studies that will be described in subsequent chapters. In these studies the SDQ-I was administered to all students in intact classrooms during regular school hours. While 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 geographically 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 children in other parts of the world, or for children who complete the instrument for other purposes or under different motivational circumstances. However, cross cultural data presented in Chapter 12 provide at least preliminary support for using this large, well-stratified norms sample for children outside Australia.

Normative data for seven of the eight individual scales and all three total scores were obtained for all 3,562 students in the normative sample. However, the General-Self scale was added more recently. For this scale only 739 responses were available, primarily from fifth grade students. Consequently, no normative comparisons are yet available for grades 2 through 4 on this one scale, and comparisons for grades 5 and 6 should be made cautiously. Note that the General-Self scale is not included in any of the total scores, so that the normative comparisons for these scores are not affected.

Nonnormalized Versus Normalized *T* Scores

Norms for the SDQ-I individual and total scales are presented as mid-interval percentiles and nonnormalized *T* scores. Norms for the control scores are presented as mid-interval percentiles and normalized *T* scores. This section presents an overview of the rationale for and underlying assumptions of the use of *T* scores.

In order to compare different scale scores, it is necessary to equate the different scales. Except in special circumstances, this requires some sort of transformation. One type of transformation (e.g., *z* scores and the nonnormalized *T* scores) equates the means and standard deviations of the different scores but does not affect the shapes of the distributions of scores. A second type of transformation (e.g., stanine scores and normalized *T* scores) forces each distribution of scores to be normally

distributed as well as have the same mean and standard deviation. Both types of transformation are based on implicit, typically untestable assumptions about the true underlying distributions of responses to different scales. The first assumes that the true, underlying mean and standard deviation of each area of self-concept are really the same and that it is only the observed means and standard deviations that differ. For example, this approach assumes that across an appropriate normative sample, children do not have systematically higher Reading self-concepts than Math self-concepts.

The use of normalized scores further assumes that the underlying distribution of self-concepts in different areas is the same and is normally distributed. Again, this assumption needs to be considered carefully. Each of the SDQ-I scales is at least moderately skewed. Intuitively, the author suggests that underlying self-concepts may actually be negatively skewed; that is, most children tend to feel positively about themselves, and only rarely does a child feel very negatively about himself or herself. The most negatively skewed of the SDQ-I scales is the Parent Relations scale. Whereas it is possible that some of this skew is due to the idiosyncratic wording of the items in the Parent Relations scale, it also appears to be the scale in which children's self-concepts are most uniformly positive. If the underlying distributions of self-concepts are positively skewed, then it may be inappropriate to normalize the response distributions.

In addition to problems in validating the implicit assumptions underlying these different transformations, there are also technical difficulties in applying them. In particular, the skews of the individual scale scores may be so extreme as to preclude the appropriate use of normalized *T* scores. Inspection of distributions of responses to the different SDQ-I scales in the normative group indicates that SDQ-I responses are negatively skewed for all eight individual scales. The skews vary from $-.41$ to -1.71 (median = $-.79$), though only two are skewed more than -1 . The skews for the three total scores vary from $-.38$ to $-.55$. While not as extremely skewed as the individual scale distributions, they are still significantly negative. On the other hand, the distributions of the control scores are relatively unskewed.

The distributions of scores on the individual scales are skewed in the direction of higher self-concept scores. To apply normalized transformations to such scores would change the essential shape of these distributions. Therefore, standardized (nonnormalized) *T* scores were generated for the individual scales by subtracting the mean and dividing by the standard deviation of raw scores for each scale and then transforming the resulting *z* scores to a mean of 50 and a standard deviation of 10 (Angoff, 1984), as is done on profiles such as that of the MMPI. These standard *T* scores are comparable in terms of standard deviation units, e.g., where $40T$ is one standard deviation below the mean, and $70T$ is one standard deviation above the mean. The user should keep in mind, however, that because the individual scales have skewed distributions, the proportion of cases falling above the mean is greater than the proportion falling below the mean. Hence, $40T$ and $70T$ are not equidistant from the mean in a percentile sense for the individual scales. The control scores, on the other hand, have distributions that more closely approximate a normal curve; therefore, normalized *T* scores were generated for the control score norms tables.

Interpretation

The comparison of different scales for the same individual child provides a special set of problems in the appropriate scaling of the SDQ-I responses. Consider, for example, the question of whether a particular child has a higher Math or Reading self-concept. The issue of appropriate scaling for this particular problem has not yet been adequately resolved. Logically, there are four approaches to making such

comparisons: the comparison of raw scale scores; the comparison of percentiles; the comparison of standard scores (e.g., *T* scores, *z* scores, factor scores) that equate the means and standard deviations for all the scales in relation to responses of a normative sample; and the comparison of standardized scores that equate the shape of the distributions — typically to a normal curve — as well as the means and standard deviations (e.g., stanine scores and normalized *T* scores). The use of these different types of comparisons is discussed below.

Raw scores. One approach is simply to compare the raw scale scores. This approach, however, is rarely defensible. For personality-type tests, minor wording changes in a few items on one scale could systematically raise or lower all scores on that scale relative to scores on other scales. (A similar caution applies to achievement test scores.) For this reason differences between two raw scale scores generally should not be used to infer underlying self-concept differences. A possible exception to this generalization, however, exists for the three academic scales of the SDQ-I.

Unlike the other SDQ-I scales and the academic scales of the other SDQ instruments, the wording of the items on the SDQ-I academic scales is strictly parallel except for the words "Reading," "Mathematics," and "All School Subjects." For this reason, it may be reasonable to compare the raw scores for just these three scales. As noted below, valid information in these scales may be lost when the scales are standardized, or normalized, or both. For example, inspection of the norms indicates that Math scale scores are systematically lower than the corresponding Reading and General-School scores and that this difference is larger for females than for males. After discussions with students who have completed these scales, the author intuitively suggests that these differences are "real" — that students generally do have lower Math self-concepts than Reading and General-School self-concepts. There is empirical support for this suggestion in that scale score comparisons are based on items with parallel wording. If, however, scores are standardized, then the mean for each scale is set at a constant value (e.g., 50 for *T* scores), and this possible "absolute" difference between Reading and Math self-concepts is lost. For this reason the raw scale scores for just the SDQ-I academic scales may be useful in addition to the *T* scores and the percentiles derived from the raw scale scores.

Percentiles. When raw scores are converted to percentiles, the child's relative position in the standardization sample can be identified. The percentile shows the percentage of children in the standardization sample whose scores fell below the child's score. Since half of the standardization sample had self-concept scores falling 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 scale is discouraged. Even in the middle range one can only assume that the child has average self-concept. Therefore, only scores at the low end of the scale are readily interpretable and diagnostically meaningful.

***T* scores.** Because the total and individual scale scores do not have highly similar distributions, comparisons between these scores should be made very cautiously. Comparisons among the three total scores are appropriate since their distributions are more similar. However, comparison of the different individual scale scores should be made cautiously because the skews for each individual scale vary substantially in some instances, particularly for the Parent Relations scale. Thus, when the *T* scores for two different scales are the same, the corresponding percentile ranks will not be the same unless the distribution of scores is also similar. Because most of the SDQ-I scales are only moderately skewed, the comparison of these specific scale scores may be reasonable.

Most SDQ-I research has focused on the first seven SDQ-I scales; consequently, these form the primary basis for the interpretation of responses to the SDQ-I. Each of these scales has high face validity. (See Figures 2 through 9, pages 5-7, for the inferred meaning of the seven SDQ-I scales and the items comprising each scale.)

In general a high scale score indicates that the child or preadolescent has a positive self-perspective in that area, whereas a low score indicates negative self-perspective. There are no absolute cut-offs for what constitutes a "high" or "low" self-concept, and interpretations should be made in relation to percentile ranks and scaled scores derived from the norms tables. Several general observations, however, may be useful. First, as mentioned earlier, half the children from the normative sample had self-concepts falling between the 25th and 75th percentiles. Self-concepts falling within this broad range indicate that a child's self-concept is neither particularly high nor particularly low. Second, particularly at the high end of the scale, relatively small response differences can result in relatively large differences in percentile ranks. For this reason, users are cautioned *not* to over-interpret differences based on responses by individual children. Third, it is expected that most children will have a range of self-concepts in different areas. Thus, even the most able children may have average or even below-average self-concepts in some areas. Similarly, even the least able children will have average or even above-average self-concepts in some areas. Finally, if a child consistently has a very high or very low self-concept across all areas, it is recommended that control scores be calculated to provide a check for inappropriate responding. In addition to these general recommendations, two special situations require further consideration.

First, a child or preadolescent may have a self-perspective that is quite unrealistic when compared to objective information. For example, a child 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 child has responded honestly, then his or her responses reflect a valid inference about self-concept even if they are unrealistic. Thus, when external indicators or the opinions of external observers differ from the child's responses, it is the external indicators that may lack validity as measures of self-concept. Nevertheless, it is recommended that control scores be calculated to test for inappropriate responding when this occurs.

Second, interpretations of the SDQ-I are based on the assumption that the child or preadolescent is responding candidly. Because the SDQ-I items and scales are straightforward, it would be easy to respond in a manner that gives either a good or a poor impression. The strength of the psychometric properties of the SDQ-I — particularly the factor analyses and relations to external criteria — suggest that positive "faking" is not usually a problem. However, SDQ-I research has nearly always been conducted in a setting in which children are assured anonymity and have little reason to respond in a distorted manner, either positively or negatively. In a setting in which respondents are externally motivated to look either 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 child or preadolescent. The first, which is recommended for use with the SDQ-I, 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. Related problems were discussed earlier in this chapter with regard to control scores used to test for inappropriate responding. Furthermore, alternative interpretations of effects of experimental interventions as inferred by self-concept responses are discussed in Chapter 10.

Total Scores

The SDQ-I contains three composite self-concept scores:

- | | |
|--------------------------------|---|
| Total Nonacademic Self-Concept | - the mean of the responses to the Physical Abilities, Physical Appearance, Peer Relations, and Parent Relations scales described above (see Chapter 1 and Figures 2 through 5) |
| Total Academic Self-Concept | - the mean of the responses to the Reading, Mathematics, and General-School scales described above (see Chapter 1 and Figures 6 through 8) |
| Total Self-Concept | - the mean of the responses to the first seven factors (1 - 7) described above |

The original justification for the total scores was based on the assumption that individual academic facets, individual nonacademic facets, and both academic and nonacademic facets are substantially correlated. However, subsequent research demonstrated that these facets are, in fact, surprisingly distinct. For example, considerable research with subjects of all ages has demonstrated that Math self-concept is nearly uncorrelated with Reading self-concept. Thus, the theoretical justification for forming a Total Academic self-concept score is weakened. Nonetheless, the SDQ-I total scores have been retained because the Total Academic score is substantially correlated with each of the academic scales (particularly General-School), the Total Nonacademic score is substantially correlated with each of the nonacademic scales, and the Total Self score is substantially correlated with the General-Self score. Furthermore, these total scores are consistent with scores from other self-concept instruments as well as the Shavelson model from which the theoretical basis of the SDQ-I is derived. Nevertheless, in their interpretation of the SDQ-I, users are encouraged to emphasize responses to the specific facets of self-concept rather than the total scores.

General-Self Scale

Historically, self-concept research has emphasized a general self-concept instead of specific facets of self. However, this general self was typically inferred on the basis of a hodgepodge of self-referent items that were not balanced with respect to any theoretically defensible components of 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 inferred a superordinate construct called general self (or sometimes called "esteem"). The most well-known scale of this sort is the Rosenberg (1965) scale that was used as the basis of the General-Self scale on SDQ instruments (also see Harter, 1982). The General-Self scale on the SDQ-I, as on Rosenberg's scale, infers 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. It refers to a student's rating of himself or herself as an effective, capable individual who is proud

of and satisfied with the way he or she is. Interpretations of general self-concept and its theoretical and empirical bases are discussed further in Chapter 4.

Individual Scales

As discussed previously, SDQ-I scales are relatively uncorrelated, and research summarized in this Manual provides strong support for the multidimensionality of self-concept based on SDQ-I responses. For these reasons, users are encouraged to use the specific SDQ-I scales instead of the total scores. Most SDQ-I research has been based on factor scores, although some studies have used raw scale scores. The focus of this research has been on comparing groups of children on separate scales, making comparisons of the same children over time, or relating the different SDQ-I scales to other criteria.

Control Scores

Interpretations of SDQ-I scores assume that the child responds appropriately when completing the SDQ-I. As indicated in Chapter 2, if there are more than four items omitted, or if other unusual problems were noted during the administration of the SDQ-I, then the responses should be interpreted cautiously if at all. Two other types of inappropriate responding are considered in this section. The first, *noncontingent responding*, refers to responses that are independent of the item content. This could take the form of either random responses, giving the same response independent of the item content, or some other nonrandom response pattern that is independent of the item content. The second, *positivity and negativity biases*, refers to the tendency of children to use the positive (agree) end of the response scale or the negative (disagree) end of the response scale independent of the item content. There are no absolute criteria for these potential problems, but six experimental control scores have been devised to provide an indication of such problems. Note, however, that no validation research, other than the normative study, has been done with these control scores. Hence, these scores should be interpreted cautiously.

The six control scores are defined according to responses from selected items and scales. The scores may be calculated by hand using the instructions in Chapter 2 and an SDQ-I Scoring and Profile Booklet. Because computation of the control scores with the hand-scoring instruction requires considerable effort, it is considered optional. However, whenever there are suspicions that a child has not responded appropriately, the computation of control scores is encouraged. Control score norms tables which provide mid-interval percentiles and normalized *T* scores based on the entire normative sample are presented in Appendix B. The definitions and interpretation of the control scores are presented below.

Control Score 1: Inconsistency of Correlated Item Pairs. This control score represents the sum of absolute (unsigned) differences in responses to 20 pairs of items. Each of the 20 item pairs was selected so that the two items came from the same SDQ-I scale, had approximately the same mean response, and were substantially correlated. Scores on this control scale vary from 0 to 80. Appropriate responding should lead to low scores, whereas inappropriate responding should lead to high scores. Only 10% of the scores in the normative sample were over 20, and only 5% were over 23. Thus, scores higher than 23 may be indicative of a noncontingent responding other than giving the same response to each item.

Control Score 2: Consistency of Uncorrelated Item Pairs. The second control score represents the sum of absolute (unsigned) differences to another 20 pairs of items. Each of the 20 item pairs was selected so that the two items came from different SDQ-I scales, had item means that differed substantially from each other,

and were nearly uncorrelated. Appropriate responding should lead to high scores whereas inappropriate responding should lead to low scores. Scores on this control scale can vary from 0 to 80 but only 10% of the scores in the normative sample were less than 11 and 5% less than 7. Scores lower than 7 may be indicative of giving the same or similar responses to all items independent of the item content.

Control Score 3: Noncontingent Summary. A third control score is defined as Control Score 2 (Consistency) minus Control Score 1 (Inconsistency). Control Score 3 has a possible range of -80 to +80. Appropriate responding should lead to high scores whereas inappropriate responding should lead to low or negative scores. Only 10% of the scores in the normative sample were less than 1, and 5% were less than -2. Scores below -2 may indicate that the responses do not depend on the content of the SDQ-I items and therefore may be invalid.

Control Score 4: Negativity Bias. The fourth control score consists of the absolute (unsigned) differences between 10 pairs of scores. In each pair of scores, one score is a response to a negatively worded item (after it has been reverse scored), and the other score is the scale raw score (based on positively worded items from the same SDQ-I scale). Because the responses to the negatively worded items are reverse scored, appropriate responding should lead to low scores whereas inappropriate responding may lead to high scores. Negativity Bias scores can vary from 0 to 40, but only 10% of the scores in the normative sample were greater than 15, and 5% were greater than 18. Scores higher than 18 suggest inappropriate responding or problems with the logical demands associated with responding to negatively worded items. As described in Chapter 13, young children often have problems responding appropriately to negatively worded items even when they are apparently responding appropriately to positively worded items. For this reason, particular care must be used in interpreting this control score. Particularly for children in grades 2 through 4, perhaps the best strategy is to use Negativity Bias to test interpretations based on other control scores.

Control Score 5: Positivity Bias. A fifth control score is the signed difference between the 10 pairs of scores used to define Negativity Bias. Positivity Bias can vary from -40 to +40, but appropriate responding should lead to scores that are close to zero. Only about 5% of the scores in the normative sample were greater than 9; such scores may indicate a positivity bias — the tendency to agree with items independent of whether the items are positively or negatively worded. Only about 5% of the scores were below -11; such scores may indicate a negativity bias — the tendency to disagree with all items whether they are positively or negatively worded. Again, because young children have problems responding appropriately to negatively worded items, this control score should be interpreted cautiously.

Control Score 6: Individual Profile Variation. The sixth control score is the standard deviation of the original seven SDQ-I scale scores calculated for an individual profile. Appropriate responding leads to moderate to high scores because the different SDQ-I scales are reasonably independent and because children are expected to have relatively higher self-concepts in some areas and relatively lower self-concepts in other areas. Very low scores suggest that either children are not responding appropriately, are not differentiating among the scales, or that they truly have self-concepts that do not vary according to the particular facet of self. Scores on Individual Profile Variation can vary from 0 to 15 or more, but only 10% of the scores in the normative sample were less than 2.7, and only 5% were less than 2.0. Scores less than 2.0 may indicate inappropriate responding. Scores on this control scale must be interpreted cautiously because a low score may accurately reflect a child's self-perceptions. Thus, Individual Profile Variation should be used to test interpretations based on other control scores.

Because these control scores have not yet been used in independent research studies, there is limited empirical support for the validity of the interpretations that are offered. Nevertheless, if control scores are extreme, interpretations of the SDQ-I should be made cautiously, if at all. It is anticipated that the control scores will be most useful for the interpretation of individual responses. An initial screening based on the control scores may also be useful for large-scale research projects as well as to insure the integrity of the data.

Discussion

Because there are as yet unresolved difficulties with the appropriate scaling of the SDQ-I scale scores, users should be cautious about making interpretations of overall profiles based on responses by an individual child. In particular, the distributions based on the Parent Relations scale and the General-Self scale are sufficiently different from the other individual scale scores and the total scores, so that the comparison of these scales with the other scales must be done cautiously. Because the other six specific scales and the three total scores are all moderately skewed in the same direction (skews of $-.38$ to $-.91$) the direct comparison of T scores and percentiles for these scales is more justified, though caution is still warranted in not over-interpreting small differences. Because of the Central Limit Theorem which states that many sample distributions, even if not normally distributed, can be approximated by a normal distribution, given a large N (Hays, 1981), the comparison of group means on these different scales is not subject to these considerations. Also, the difficulties in the interpretations of profile patterns for different SDQ-I scales do not introduce new problems for the interpretation of each scale individually or the interpretation of changes over time for individual subjects.

Chapter 4. Multidimensional Assessment of Self-Concept

The multidimensional structure of self-concept is a basic assumption of the Shavelson model and is very important to understanding self-concept. For example, a child may have a *positive* self-concept with regard to peer relations but a *poor* self-concept with regard to mathematics. This chapter provides the theoretical basis and empirical support for the multidimensional structure of self-concept. In addition, data are presented illustrating that the SDQ-I is a valid measure of the multidimensional model.

The development of the SDQ-I began with the Shavelson model described below. Item pools derived from the Shavelson model were developed, and preliminary analyses were used to refine the items designed to measure each scale. Once a suitable version of the SDQ-I had been developed, additional research was conducted to examine characteristics of the SDQ-I and test assumptions of the Shavelson model. This research led to refinement of both the SDQ-I and the Shavelson model. Thus, the development of the SDQ-I represents a dynamic interplay between theory and empirical research. Such an interchange is the goal of construct validation (Cronbach & Meehl, 1955).

The Shavelson Model

Reviews of self-concept research consistently identify shortcomings such as the lack of a theoretical basis for defining and interpreting the construct and the poor quality of instruments used to measure it. In an attempt to remedy this situation, Shavelson, Hubner, and Stanton (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 especially influenced by evaluations by significant others, reinforcements, and attributions for one's own behavior.

In the Shavelson model, self-concept is further defined by seven major features:

- It is organized or structured, in that people categorize the vast amount of information they have about themselves and relate these categories to one another.
- It is multifaceted, and the particular facets reflect a self-referent category system adopted by a particular individual and/or shared by a group.
- It is hierarchical with perceptions of personal behavior at the base moving to inferences about self in superordinate areas (e.g., academic — English, science, history, mathematics), and then to inferences about oneself in general.

- The hierarchical general self-concept — the apex of the model — is stable, but as one descends the hierarchy, self-concept becomes increasingly situation-specific and, as a consequence, less stable.
- Self-concept becomes increasingly multifaceted as the individual moves from infancy to adulthood.
- It has both a descriptive and an evaluative aspect; individuals may describe themselves ("I am happy") and evaluate themselves ("I do well in mathematics").
- It can be differentiated from other constructs such as academic achievement.

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 broken into self-concepts in particular subject areas (e.g., mathematics, English). Nonacademic self-concept is broken into three areas: social self-concept (which is broken into relations with peers and with significant others), emotional self-concept, and physical self-concept (which is broken into physical abilities 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 of limited generality, quite specific, and very closely related to actual behavior. The Shavelson model posits a structure of self-concept that resembles British psychologists' hierarchical model of intellectual abilities in that general self-concept (like Spearman's "g") is at the apex, and it can be divided into two components which are then divided into group and specific factors.

The self-concept facets proposed in the Shavelson 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. Studies employing five commonly used instruments provided modest support for the separation of self-concept into social, physical, and academic facets. However, these three facets were not clearly identified in any one of the instruments. Shavelson et al. were unable to identify any existing instrument which would measure multiple facets of self-concept as posited in the model, nor was the multifaceted nature of self-concept proposed by Shavelson et al. widely investigated by other researchers.

Some researchers (e.g., Coopersmith, 1967; Marx & Winne, 1978) argued that the facets of self-concept were so heavily dominated by a general factor that they could not be adequately differentiated. Coopersmith, on the basis of preliminary research with his *Self-Esteem Inventory*, argued that "preadolescent children make little distinction about their worthiness in different areas of experience or, if such distinctions are made, they are made within the context of the over-all, general appraisal of worthiness that children have already made" (page 6). Despite this assertion, factor analyses of responses to the Coopersmith instrument do reveal multiple factors, though these factors are not readily interpretable and bear little resemblance to those that the instrument was designed to measure (Marsh & Smith, 1982). Marx and Winne classified the subscales from three commonly used self-concept instruments into the academic, social, and physical facets hypothesized by Shavelson et al. and used multitrait-multimethod analyses to compare responses from different instruments. They found that responses to each of the three facets demonstrated some agreement across instruments (convergence), but responses to the different scales could not be adequately differentiated (divergence). They concluded that "self-concept seems more of a unitary concept than one broken into distinct subparts or facets" (page 900). Shavelson and Bolus (1982) argued that there was insufficient justification for Marx and Winne's classification of subscales into facets. Reanalyzing the Marx and Winne data by taking a single scale from each

instrument which best represented each of the three facets, they were able to demonstrate modest divergent validity.

In contrast to claims that self-concept is largely unidimensional, most studies which systematically examined the multidimensionality of self-concept found support for the multifaceted interpretation (e.g., Fernandes, Michaels, & Smith, 1978; Fleming & Watts, 1980; Harter, 1982; Marsh, Parker, & Smith, 1983; Marsh, Relich, & Smith, 1983; Marsh, Smith, & Barnes, 1983; Michaels, Smith, & Michaels, 1975; Piers & Harris, 1964; Shavelson & Bolus, 1982; Shepard, 1979; Wylie, 1979). Soares and Soares (1977, 1982), for example, described a theoretical conceptualization of self-concept which is even more extreme than the Shavelson model. They argued that the low correlations observed among different areas of self-concept suggest a model of nearly independent factors of self-concept instead of the strong hierarchical ordering posited by Shavelson et al. Nevertheless, until recently, factor analyses of the most commonly used self-concept instruments typically failed to identify the scales that the instrument was designed to measure (cf. Marsh & Smith, 1982; Shavelson et al., 1976). Byrne (1984) noted that "Many consider this inability to attain discriminant validity among the dimensions of SC [self-concept] to be one of the major complexities facing SC [self-concept] researchers today" (pages 449-450).

Thus, conflicting interpretations fall along a continuum where the representation of self-concept varies from a unidimensional construct to one of multiple facets that are nearly independent. The hierarchical representation of self-concept in the Shavelson model may be viewed as consistent with either extreme, depending on the strength of the hierarchy. However, when the structure of the hierarchy is so strong that facets can reasonably be represented as a single factor, or so weak that the facets are nearly independent, then the usefulness of the hierarchical representation becomes dubious.

Through the mid-1970s, self-concept instruments typically consisted of a hodgepodge of self-referent items, and "blind" applications of exploratory factor analyses 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. Factor analysis is then used to test these a priori facets (e.g., Boersma & Chapman, 1979; Dusek & Flaherty, 1981; Fleming & Courtney, 1984; Harter, 1982; Soares & Soares, 1982; and particularly SDQ-I research summarized in this Manual). Recent reviews of this research (Byrne, 1984; Marsh & Shavelson, 1985; Shavelson & Marsh, 1986) make obvious the multifaceted structure of self-concept. They indicate that self-concept cannot be adequately understood if its multidimensional character is ignored.

Even those who accept the multifaceted nature of self-concept do not agree on the identity of the specific dimensions that comprise self-concept and how they are structured. One purpose of the SDQ-I was to provide a reliable and valid instrument to test the assumptions underlying the conceptual structure of self-concept posited by Shavelson. Byrne (1984), in her review of self-concept models, concluded that "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" (page 449).

Development and Factor Structure of the SDQ-I

The SDQ-I, based on the Shavelson model, was originally developed to measure four areas of nonacademic self-concept (Peer Relations, Parent Relations, Physical Abilities, and Physical Appearance) and three areas of academic self-concept

(Reading, Math, and General-School). Emotional self-concept, though hypothesized by Shavelson, was excluded since preliminary investigations suggested that young children had difficulty with these items, and a satisfactory scale could not be constructed. (This scale is included, however, on SDQ-II and SDQ-III.) Extensive item pools were developed to measure each of the proposed factors. Items were adapted from previous self-concept research or developed to measure components of self-concept hypothesized in the Shavelson model. Items were tried out, and results were used to refine subsequent versions of the instrument. Criteria used for the inclusion of items on subsequent drafts included the results of item- and factor-analyses. In addition, items were rewritten or excluded if they were difficult for children and adolescents to comprehend or respond to. On the basis of this preliminary research, the seven-factor version of the SDQ-I was formulated (see Marsh, Parker, & Smith, 1983).

The seven originally hypothesized factors have been replicated across more than a dozen factor analytic studies of responses to the SDQ-I by diverse populations of children and preadolescents (e.g., Marsh, 1986a, 1987b; Marsh, Barnes, Cairns, & Tidman, 1984; Marsh, Relich, & Smith, 1983; Marsh & Smith, 1987; Marsh, Smith, & Barnes, 1983, 1984, 1985).

Exploratory Factor Analyses

This factor structure was tested with a single summary factor analysis performed on all 3,562 sets of responses in the normative data base. Responses to positively worded items from the original seven SDQ-I factors were used. Consistent with earlier SDQ-I research, the eight positively worded items from each of the seven SDQ-I scales were divided into four 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. The factor analysis was performed on responses to these 28 item pairs (for further discussion and rationale, see Marsh, Barnes, Cairns, & Tidman, 1984; Marsh & O'Niell, 1984) using a commercially available SPSS program (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975), with iterated communality estimates, a Kaiser normalization, and an oblique rotation to a final solution with delta set to -2.0.

Results clearly identified each of the SDQ-I factors (see Table 1). Factor loadings for variables designed to measure each factor (target loadings are in boldface) were substantial, ranging from .46 to .85 (median = .73). The nontarget loadings were much smaller, ranging from -.02 to .19 (median = .03). Correlations among the factors were modest, ranging from .03 to .47 (median = .12), and were much smaller than the coefficient alpha estimates of reliability which range from .80 to .91 (see Chapter 5). The largest correlations occurred among the first three nonacademic factors and between the General-School and the other two academic self-concepts. These results are consistent with a hierarchical ordering as hypothesized by Shavelson et al. Despite the moderate correlation between General-School and Reading (.34) and between General-School and Math (.47), the correlation between Reading and Math self-concepts (.05) was close to zero. The near-zero correlation between Math and Reading self-concepts is consistent with the Internal/External Frame of Reference Model described in Chapter 7. These results provide strong support for the multidimensionality of self-concept and particularly for the facets hypothesized in the Shavelson model.

The Shavelson model not only posits that self-concept is multifaceted but also that it is hierarchically ordered and becomes more distinct with age. These assumptions were tested in a series of exploratory factor analyses, conducted separately by grade, on a sample of 658 students in grades 2 to 5 (for methodological details see Marsh, Barnes, Cairns, & Tidman, 1984; also see Marsh & Hocevar, 1985; Marsh &

Table 1. Exploratory Factor Analysis for the Total Normative Sample (N = 3,562)

Variable	Factor							% Specific Variance Plus Error
	Physical	Appearance	Peers	Parents	Reading	Math	School	
Physical 1	.67	.03	.07	.01	.00	.01	.01	45.3
Physical 2	.56	.14	.07	.06	-.01	.09	-.02	41.3
Physical 3	.84	.05	.06	.01	.02	.02	.03	64.3
Physical 4	.75	.03	.12	.03	.00	-.01	.07	58.9
Appearance 1	.04	.78	.02	.08	.02	.07	.01	61.8
Appearance 2	.02	.80	.06	.03	.04	.00	.06	65.3
Appearance 3	.13	.67	.19	-.02	.01	.00	.07	63.8
Appearance 4	.10	.64	.18	.04	.01	-.01	.08	59.6
Peers 1	.09	.00	.64	.09	.00	.01	.02	43.1
Peers 2	.04	.14	.63	.06	.03	.01	.10	50.5
Peers 3	.06	.08	.68	.02	.04	.07	.00	49.1
Peers 4	.12	.17	.63	.05	.00	.01	.06	55.5
Parents 1	.05	.04	.04	.57	.01	.01	.04	31.9
Parents 2	.02	.02	.01	.56	.06	.04	.04	30.2
Parents 3	.02	.04	.11	.72	.04	.03	.03	50.1
Parents 4	.00	.04	.07	.78	.01	-.01	.04	50.9
Reading 1	.00	.02	.00	.01	.78	-.02	.09	60.6
Reading 2	.00	.01	.02	.02	.85	.02	.07	68.2
Reading 3	.02	.04	.03	.06	.76	.00	.14	64.7
Reading 4	.00	.02	.03	.06	.76	-.01	.14	64.4
Math 1	.03	.04	.02	.02	-.02	.75	.17	69.2
Math 2	.03	.02	.03	.04	.01	.78	.17	73.7
Math 3	.03	.01	.03	.05	.00	.79	.17	74.9
Math 4	.03	.03	.04	.00	.01	.81	.14	75.7
School 1	-.01	.07	.04	.01	.08	.10	.65	51.0
School 2	.06	.11	.10	-.01	.12	.17	.46	47.6
School 3	.01	-.02	.00	.09	.09	.16	.65	59.3
School 4	.05	.01	.04	.03	.09	.12	.74	64.2
Correlations Among Factors								
Physical								
Appearance	.26							
Peers	.32	.37						
Parents	.12	.14	.22					
Reading	.03	.09	.03	.13				
Math	.11	.11	.13	.10	.05			
School	.11	.19	.20	.16	.34	.47		

Note: 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. Decimal points for factor loadings and correlations are omitted.

Shavelson, 1985). These analyses examined the first-order factor structures underlying responses to items on the SDQ-I at each grade level separately, the consistency of this structure across grade levels, and the hierarchical structure of these first-order factors. Based on previous research with the SDQ-I and the self-concept model, it was predicted that: (a) the seven SDQ-I factors would be identified at each grade level; (b) the factor loadings would be similar at each grade level; (c) the size of correlations among the oblique factors would become smaller with age; and (d) at each grade level, the hierarchical structure among the facets would resemble the structure hypothesized by Shavelson et al.

The same four item pairs were formed from the eight items designed to measure each of the seven facets of self-concept as described above. Responses to these 28 item pairs were factor analyzed separately at each grade level with conventional exploratory factor analytic procedures. Results indicate that the seven facets (factors) were identified at each of the four grade levels and that factor loadings for each item pair (target loadings are in boldface) were consistently high on the factor it was designed to measure and low on other factors (see Table 2), providing support for the first two predictions. The factors were particularly well defined for grades 4 and 5, whereas the separation among the academic factors was less clear for grades 2 and 3.

The pattern of correlations among the seven oblique factors (see Table 2) offered support for hypothesis (d). At each grade level the highest correlations tended to occur between the General-School factor and the other academic dimensions and among the scales measuring nonacademic dimensions.

Despite the similarity in this pattern of correlations, there was a consistent decline with age in the size of correlations among the factors. In support of prediction (c), the median correlations among the factors decreased with increasing grade level (.27, .19, .18, and .14 in grades 2 through 5, respectively; see Marsh, Barnes, Cairns, & Tidman, 1984). This trend occurred despite the fact that the reliabilities (coefficient alphas are in bold italics) for the SDQ-I factors were somewhat more reliable for the older children. The Marsh et al. findings strongly support the assumptions of a multifaceted, hierarchical self-concept. In addition, they support the validity of interpretations based on SDQ-I scales by indicating that the scales measure distinct factors that are related in theoretically defined ways.

Confirmatory Factor Analyses

Though the exploratory factor analytic technique employed by Marsh et al. is widely used for initial investigation, it is not entirely suitable for testing assumptions of the Shavelson model. It does not allow a particular structure to be statistically tested against alternative structures, and the structure of the model cannot be controlled beyond setting the number of factors and perhaps their level of correlation. Moreover, it offers no way to examine the structure of first-order and second-order factors within the same analysis. These limitations are overcome with the use of confirmatory factor analysis. The results of such analyses, using LISREL V (Joreskog & Sorbom, 1981), were examined in subsequent studies (Marsh & Hocevar, 1985; Marsh & Shavelson, 1985). The confirmatory factor analytic studies are summarized below.

First-order factor structure. A confirmatory factor analysis was used to test the first-order structure of responses to the seven-scale SDQ-I and its invariance across grade levels. It was hypothesized that: (a) the responses to the SDQ-I could be explained by seven factors; (b) each item pair, as employed in the previous studies, would load only on the factor it was designed to measure; (c) the seven factors would be correlated; and (d) the error and uniqueness terms for the measured variables

Table 2. Exploratory Factor Analysis by Grade Level

Variable	Factor						
	Physical	Appearance	Peers	Parents	Reading	Math	School
Factor Loadings, Grade 2 (n = 170)							
Physical 1	.50	.05	.06	.08	-.06	.14	.09
Physical 2	.62	.02	-.04	.22	.04	-.03	.09
Physical 3	.64	.10	-.05	.13	-.08	.14	.16
Physical 4	.69	-.10	.22	.05	.15	-.01	-.11
Appearance 1	.01	.67	.03	.16	-.01	.06	.02
Appearance 2	-.15	.70	.04	.18	.04	.01	-.08
Appearance 3	.19	.47	.11	.03	-.01	.36	-.26
Appearance 4	.00	.55	.00	.21	.09	.32	-.13
Peers 1	.05	-.01	.60	.16	-.11	.12	-.11
Peers 2	-.16	-.12	.47	.39	.20	.28	-.13
Peers 3	.13	.06	.27	.20	-.05	.45	-.12
Peers 4	.25	.10	.39	.16	.15	.29	-.29
Parents 1	.29	.22	.21	.34	-.08	-.16	.13
Parents 2	.18	-.06	.05	.34	.29	-.02	.13
Parents 3	.01	.17	.12	.49	.16	.05	.07
Parents 4	.15	.13	-.04	.52	.39	.01	-.05
Reading 1	.05	.22	.34	-.16	.49	-.13	.20
Reading 2	.19	.28	.27	-.09	.48	-.04	.02
Reading 3	.07	.23	.38	.10	.40	-.10	.12
Reading 4	.08	.07	.00	.18	.56	.11	.05
Math 1	.11	.10	.12	.00	-.08	.51	.33
Math 2	.08	-.05	.12	.11	.32	.50	.39
Math 3	-.02	.10	.16	.24	.00	.39	.59
Math 4	.14	-.06	.06	-.07	.20	.58	.32
School 1	.04	.20	.46	.01	-.06	-.11	.49
School 2	.14	.17	.07	.01	.21	.17	.36
School 3	-.09	.05	.09	-.02	.46	.19	.35
School 4	.14	.21	.12	-.24	.29	.21	.41
Correlations Among Factors							
Physical	.78						
Appearance	.18	.85					
Peers	.29	.34	.83				
Parents	.34	.32	.36	.80			
Reading	.19	.26	.31	.28	.84		
Math	.26	.27	.28	.27	.24	.85	
School	.20	.12	.15	.00	.33	.23	.82

Note: 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. Decimal points for factor loadings and correlations are omitted. Values in bold italics are coefficient alphas, which provide an estimate of internal consistency for each of the factors.

Table 2. (continued) Exploratory Factor Analysis by Grade Level

Variable	Factor				
	Physical	Appearance	Peers	Parents	School
Factor Loadings, Grade 3 (n = 103)					
Physical 1	31	04	08	30	
Physical 2	70	14	-12	-07	-24
Physical 3	58	09	20	07	-05
Physical 4	48	14	29	01	-05
Appearance 1	-10	73	-02	05	01
Appearance 2	09	69	-15	01	10
Appearance 3	06	59	22	04	12
Appearance 4	17	65	24	-07	08
Peers 1	39	-05	08	14	02
Peers 2	27	04	42	-26	-08
Peers 3	18	-14	42	00	-10
Peers 4	07	16	74	-03	-06
Parents 1	14	20	-10	69	35
Parents 2	04	-06	21	37	-35
Parents 3	-04	-04	16	49	40
Parents 4	06	13	17	20	14
Reading 1	-13	08	03	04	12
Reading 2	-18	09	-05	44	09
Reading 3	18	06	-15	28	02
Reading 4	-03	02	18	35	13
Math 1	14	05	-10	-07	53
Math 2	05	08	06	-03	73
Math 3	-09	-01	37	27	57
Math 4	-23	10	44	08	66
School 1	04	25	-09	-03	43
School 2	24	00	-03	10	59
School 3	02	07	-02	-03	27
School 4	-13	17	16	13	31
Correlations Among Factors					
Physical	71				
Appearance	17	81			
Peers	28	15	72		
Parents	11	13	18	66	
Reading	09	20	21	17	74
Math	09	19	19	19	22
School	-12	14	12	26	19

Note: 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. Decimal points for factor loadings and correlations are omitted. Values in bold italics are coefficient alphas, which provide an estimate of internal consistency for each of the factors.

Table 2. (continued) Exploratory Factor Analysis by Grade Level

Variable	Factor				
	Physical	Appearance	Peers	Parents	School
Factor Loadings, Grade 3 (n = 134)					
Physical 1	77	-01	08	10	
Physical 2	50	12	08	-05	02
Physical 3	68	08	08	05	14
Physical 4	62	16	11	06	01
Appearance 1	08	77	03	17	07
Appearance 2	12	77	-03	08	00
Appearance 3	18	60	27	-04	07
Appearance 4	-03	57	38	-05	03
Peers 1	20	-05	66	11	04
Peers 2	14	32	41	02	00
Peers 3	02	13	69	03	00
Peers 4	13	21	63	07	00
Parents 1	10	02	-11	62	-08
Parents 2	07	03	13	45	18
Parents 3	11	00	22	75	01
Parents 4	-10	13	08	71	-07
Reading 1	-04	05	-11	00	85
Reading 2	-02	05	11	00	75
Reading 3	08	06	01	14	53
Reading 4	00	02	18	04	76
Math 1	11	07	-02	01	66
Math 2	14	-04	02	03	69
Math 3	-01	04	02	02	82
Math 4	13	01	01	02	85
School 1	01	07	-04	06	10
School 2	01	15	15	04	12
School 3	01	-07	-03	04	20
School 4	05	-06	07	03	48
Correlations Among Factors					
Physical	80				
Appearance	29	90			
Peers	35	43	87		
Parents	19	18	20	77	
Reading	03	12	15	14	85
Math	19	03	08	09	19
School	18	22	18	13	18

Note: 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. Decimal points for factor loadings and correlations are omitted. Values in bold italics are coefficient alphas, which provide an estimate of internal consistency for each of the factors.

Table 2. (continued) Exploratory Factor Analysis by Grade Level

Variable	Factor						
	Physical	Appearance	Peers	Parents	Reading	Math	School
Factor Loadings, Grade 3 (n = 251)							
Physical 1	65	00	08	-02	05	-05	-08
Physical 2	44	23	-06	08	-06	23	09
Physical 3	77	04	12	01	-01	10	06
Physical 4	83	00	04	04	02	-03	04
Appearance 1	00	80	04	09	02	10	-05
Appearance 2	-05	80	05	10	06	02	-05
Appearance 3	12	66	19	-08	-03	-07	14
Appearance 4	21	55	19	04	05	-09	14
Peers 1	09	01	61	02	13	-03	-05
Peers 2	02	19	49	06	-01	10	07
Peers 3	09	09	62	14	-07	-14	32
Peers 4	07	11	66	04	01	10	-02
Parents 1	05	08	06	50	09	04	-11
Parents 2	-10	-10	25	49	11	09	-10
Parents 3	10	05	01	75	01	-06	14
Parents 4	-02	09	-04	76	01	-02	17
Reading 1	-05	06	03	06	78	03	10
Reading 2	02	04	-05	13	76	-14	21
Reading 3	02	08	11	04	79	04	11
Reading 4	06	-07	06	06	77	10	06
Math 1	00	03	-08	05	00	79	11
Math 2	08	02	13	04	07	72	16
Math 3	01	01	04	-04	02	76	28
Math 4	03	01	07	-01	00	76	24
School 1	-09	03	07	-04	12	12	65
School 2	16	10	05	-05	16	23	38
School 3	-05	-05	00	19	20	25	52
School 4	07	-04	02	10	19	12	64
Correlations Among Factors							
Physical	78						
Appearance	23	87					
Peers	22	30	81				
Parents	06	15	23	79			
Reading	05	05	16	24	90		
Math	10	08	09	11	12	91	
School	13	15	17	13	34	43	85

Note. 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. Decimal points for factor loadings and correlations are omitted. Values in bold italics are coefficient alphas, which provide an estimate of internal consistency for each of the factors.

would be uncorrelated. This model differed from that in the previous analyses primarily in that all nontarget loadings were constrained to be zero, thus representing a much more restrictive, "simple structure" solution. The goodness of fit of the seven-factor model was examined for each grade level separately. Then various parameters were required to be constant across grade levels as tests of the invariance of the factor structure across these grade levels. Results demonstrated that the goodness of fit indicators suggest a good fit to the data for each grade level when considered separately and across all four grade levels when the factor loadings were allowed to vary independently for each grade level. These results demonstrate that the first-order factor structure based on the self-concept model was supported at each grade level.

A second, more restrictive model was then tested in which the factor loadings were constrained to be the same at each grade level. This constraint is normally considered to be the minimal condition for factorial invariance (see Alwin & Jackson, 1981; Marsh, 1985b; Marsh & Hocevar, 1985). The goodness of fit for this model provided strong evidence that the factor structure was invariant across the four age groups. In a third model, the factor loadings and the correlations among the factors were constrained to be the same across the four age levels, but the goodness of fit measures for this model indicated that it did not fit the data as well as the model in which factor correlations were allowed to differ for each grade level.

These results support both the hypothesized factor structure of the SDQ-I and the Shavelson model upon which it was based. The factors were well defined in that every factor loading at each grade level was large, statistically significant, and reasonably invariant over grades. However, the size of factor correlations did vary across the grade levels, and the direction of this variation was consistent with the earlier conclusion that the factors become less correlated and more distinct with age.

Higher order factor structure. Correlations among the self-concept factors were estimated in the analyses described above, but no special assumptions about the pattern of correlations were made. However, both the Shavelson model and the design of the SDQ-I assume that there is a systematic hierarchical ordering of the self-concept facets which underlie the correlations among first-order factors. For example, one reasonable hypothesis would be that the seven first-order factors would form two second-order factors for the nonacademic and academic facets of self-concept, a finding which would be consistent with the Shavelson model. Results of previous research and the correlations among the factors in the Marsh et al. study suggest several complications for this model. First, the Parent Relations factor was as highly correlated with some of the academic factors as with the nonacademic factors. Second, although the Math and Reading factors were each substantially correlated with the General-School factor, they were not substantially correlated with each other. These data suggest that the higher order structure underlying the SDQ-I factors may be more complicated than was previously assumed.

Higher order factor models were therefore developed to test alternative configurations of the higher order factor structure underlying the first-order factors. In each of the higher order factor models, the seven SDQ-I factors were defined as before. However, unlike previous analyses, the factor correlations were explained in terms of a higher (second) order factor self-concept structure. That is, the second-order factors were posited to explain the covariation among the first-order factors. Conceptually, this is as if a second factor analysis were performed on the matrix of correlations among the first-order factors, though the first- and second-order analyses are actually performed simultaneously in the LISREL approach to higher order factor analysis. Because hierarchical models are more parsimonious than first-order models (i.e., are based on fewer parameter estimates), they cannot fit the

data any better than first-order models. If however, a hierarchical model fits the data nearly as well as a corresponding first-order model, then there is support for the first-order model (see Marsh, Balla, & McDonald, 1988; Marsh & Hocevar, 1985, for further discussion of goodness of fit).

Four competing models were tested (see Figure 11). The chi-square fit statistics for each model are shown at the bottom of Figure 11. Poor fit to the model is indicated if the chi-square value greatly exceeds the degrees of freedom shown in parentheses. For Model 1, a single, general self-concept factor was proposed to explain the relationships among the seven first-order factors. The chi-square of 432 greatly exceeds the 56 degrees of freedom. Thus the model did not provide an adequate fit at any of the grade levels and was therefore rejected. In Model 2, two second-order factors were proposed — one defined by the four nonacademic factors and one defined by the three academic factors. It fit the data better than Model 1, but not nearly as well as the first-order model. In Model 3, the observation that the Parent Relations factor was related to both academic and nonacademic factors was taken into account by allowing this factor to load on both of the second-order factors of Model 2. Models 2 and 3 differed only in this "dual" loading of the Parent Relations factor, but the improvement (reduction in chi-square value) was statistically significant and supported the earlier observations that the Parent Relations factor is related to both nonacademic and academic self-concepts.

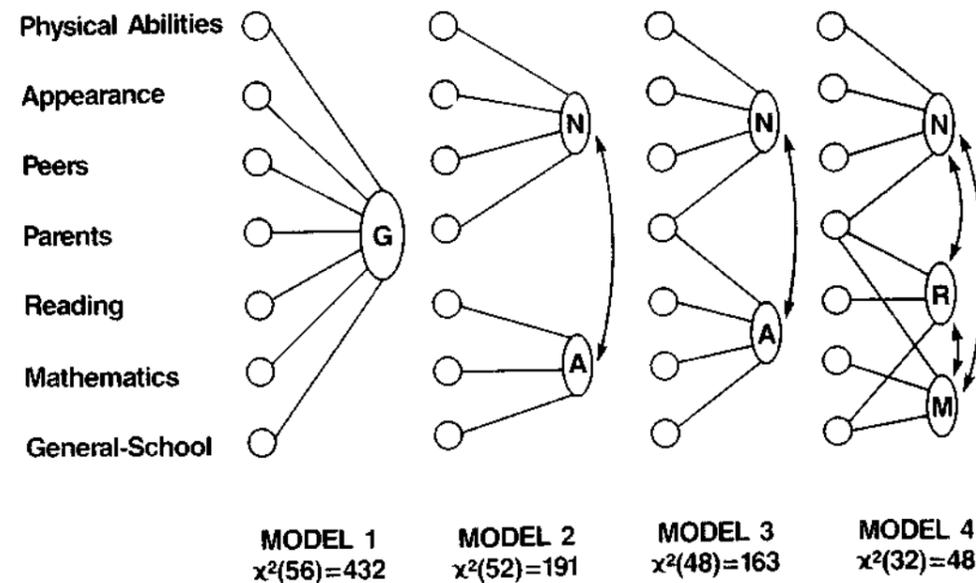


Figure 11. Higher Order Factor Structure of the SDQ-I

From Marsh, H.W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, 26, p. 114. Copyright 1985 by the American Psychological Association. Reprinted by permission of the publisher.

Model 4 took into account the previous finding that SDQ-I self-concepts in Reading and Math are nearly uncorrelated. Two second-order academic self-concept factors — Reading/Academic and Math/Academic — and a second-order nonacademic self-concept factor were posited. As in Model 3, the Parent Relations factor was allowed to load on each of the two second-order academic factors and on the nonacademic factor, as was the General-School factor. It was also proposed that the

three second-order factors were correlated, which is mathematically equivalent to saying that they combine to form a third-order, general self-concept. Results indicate that Model 4 fit the data significantly better than any of the other second-order models (Models 1-3), and nearly as well as the first-order factor model examined in the previous section.

In summary, four competing models were proposed to explain the hierarchical structure of the seven SDQ-I factors. In separate analyses of data at each grade level, the same model (Model 4 in Figure 11) provided the most accurate description of the hierarchical structure. This model is consistent with the Shavelson et al. assumption that self-concept is hierarchically ordered, but the particular form of this higher order structure is more complicated than they proposed. In particular, there seems to be a clear separation of Reading and Math self-concepts so that they cannot be incorporated into a general academic self-concept. The theoretical basis of this separation of Reading and Math self-concepts is explored further in Chapter 7.

Hierarchical structure of SDQ-III. Subsequent research (Marsh, 1987c) explored the hierarchical structure of the SDQ-III, an instrument for late-adolescents and young adults. As in the SDQ-I research just summarized, a first-order model fit the data well. Simple hierarchical models like Models 1 and 2 (in Figure 11) were unable to provide an adequate fit to the data. On the academic side of the hierarchy, two second-order factors — Math/Academic and Verbal/Academic — were clearly supported, but the structure of the hierarchy was not so clear on the nonacademic side. In particular, the second-order Physical and Social factors were not clearly differentiated even though each of the first-order factors that comprise these second-order factors was. Physical Abilities and Physical Appearance were substantially correlated, but correlations between Physical Appearance and Opposite Sex Relationships and between Physical Abilities and Same Sex Relationships were as high or higher. Thus, in the best fitting model, all the physical and social factors were incorporated into a second-order Physical/Social factor. Though these results appear to contradict the Shavelson model, it should be noted that the best fitting model for the SDQ-I in Figure 11 also combines the physical and social factors into a single second-order factor.

In Search of a General Self-Concept

The emphasis of SDQ-I research has been on the multifaceted nature of self-concept, the measurement of distinguishable facets of self-concept, and the relationships between these specific facets and a wide array of specific external criteria. In support of the Shavelson model, research summarized in this Manual clearly demonstrates that self-concept is multifaceted and cannot be adequately understood if this multidimensionality is ignored. However, there is also need for theoretical and empirical research to justify and define overall or general self-concept. Though such a construct is widely inferred, it is typically ill-defined and is probably the basis of much confusion in self-concept research. The purpose of this section is to review alternative uses of the term *general self-concept*, relate these to the SDQ-I, and describe preliminary research with the SDQ-I.

Most investigators regard self-concept as a unidimensional construct and represent it with a single score called "overall," "total," or "general self-concept." However, if one examines these uses carefully, there are clearly many different definitions of what constitutes general self-concept (see Marsh, 1986b; Marsh & Shavelson, 1985). The most common, an agglomerate self-concept, is a total score for a broad cluster of diverse self-report items which lack a coherent focus. Here, the construct is vaguely defined, and there is little rationale for the potpourri of items that are used

to measure it. Many commonly used instruments attempt to measure a diverse set of facets, but the different facets have not been empirically verified nor their contribution balanced. Instead, responses are simply summed to form a total score that is taken to be a measure of general self-concept. Such a construct cannot be adequately characterized and is idiosyncratic to particular instruments. For example, Marsh and Smith (1982) compared responses to the *Coopersmith Self-Esteem Inventory* (Coopersmith, 1967) and the *Sears Self-Concept Inventory* (Sears, 1963). Inspection of the content of each instrument suggested only a modest overlap in the aspects of self-concept tapped. Therefore it was not surprising that even total scores from the two instruments correlated only .42. This agglomerate use of general self-concept is particularly dubious and probably led to many of the contradictory findings which abound in self-concept research. Its use might be justified when the content of the total score can be characterized and the score is balanced with respect to facets derived from a theoretical model and empirically identified through factor analyses, as with the Total Self score on the SDQ-I. Even here, however, considerable information is lost by averaging across scores representing reasonably independent facets of self-concept. If only an agglomerate self is considered, the multidimensionality of self-concept is ignored.

A second, more justifiable use of the term general self-concept involves scales that are specifically designed to measure a relatively unidimensional construct that is superordinate to specific facets of self-concept. Items in such scales do not refer to self-concept in particular facets but rather infer a general sense of self-worth or self-competence that could be applied to different areas. This is the approach employed by Rosenberg (1965, 1979), Harter (1982, 1983), and the General-Self scales on the SDQ instruments developed for older subjects. Examples of items from the General-Self scale of the SDQ-III are "Overall, I have pretty positive feelings about myself," "Overall, nothing I do is very important," and "Overall, I have a lot of self-confidence." Factor analytic studies reported elsewhere with the SDQ-II, SDQ-III, and the Perceived Competence Scale for Children (Harter, 1982) demonstrate that such a factor can be clearly identified and is distinguishable from other dimensions of self-concept. Particularly for responses by late-adolescents on the SDQ-III, the responses to the General-Self scale are internally consistent and have surprisingly low correlations with other, more specific facets of self. Nevertheless, there has been little theoretical or empirical research on how this superordinate self-concept relates to more specific facets of self-concept, or on its usefulness in combination with specific facets of self-concept for predicting external validity criteria.

A third definition, a higher order self-concept, refers to an inferred construct which is not directly measured. The general self-concept that appears at the apex of the Shavelson model and the general factor in the second-order factor analyses of the SDQ-I are examples of this third use.¹

The use of a construct of general self-concept as the apex of a hierarchy, as in the Shavelson model, has important theoretical implications. Unlike the other two uses, this general self-concept cannot be tied to a specific set of items but is an unobserved construct that is itself defined by unobserved constructs (i.e., it is a higher order factor). As with the total score, it represents some average of specific facets of self-concept, and its breadth is limited by the scope of specific scales that are included in the analysis. Thus, in a study that examined only different areas of academic self-concept, such a general self-concept would necessarily be limited to a general academic self-concept. As with the General-Self scale, the apex in Figure 11

¹Marsh (1986b) also described a weighted average general self-concept in which scores for specific facets are weighted by their relative importance and a discrepancy model in which general self is defined as a function of differences between actual and ideal self-concepts. However, empirical support for the latter conceptualizations is weak, and they are not considered further in this Manual.

implies a self-concept that is superordinate to the specific self-concept facets. Additional research, using second-order factor analyses performed on responses from instruments that include a well-defined general-self scale in addition to more specific scales, would help in understanding the construct.

The Marsh (1987c) study, based on the SDQ-III, is particularly relevant to this discussion of general self-concept. In that study Marsh explored the hierarchical structure of the SDQ-III. The SDQ-III contains a relatively unidimensional, superordinate scale called General-Self. In the hierarchical model tested in this study, the General-Self scale was hypothesized to contribute directly to the hierarchical general self-concept that appeared at the apex of the hierarchy. There was empirical support for this structure. Also, correlations between the General-Self scale and the hierarchical self were consistently close to .90. This suggests that the General-Self scale and the hierarchical general self-concept are very similar when inferred on the basis of responses to the SDQ-III.

Addition of a General-Self Scale to the SDQ-I

As already noted the original version of the SDQ-I did not contain a General-Self scale. This exclusion was perhaps justified by the emphasis on a multidimensional self-concept in SDQ research. For three different reasons, however, this original version of the SDQ-I was revised to include a General-Self scale. First, empirical findings based on both the SDQ-II and the SDQ-III instruments and findings by other researchers indicated that a relatively unidimensional, superordinate scale did exist. Second, theoretical research such as described by Marsh (1986b; see above) indicated the need for such a scale. Third, this type of scale has been the basis of a considerable amount of self-concept research. Thus the inclusion of a General-Self scale on the SDQ-I provides a better basis of comparison between SDQ research and other self-concept research.

In the first large-scale application of the revised SDQ-I instrument, Marsh, Smith, and Barnes (1985) examined responses from 559 fifth-grade students. Responses to the eight positively worded items which comprise the General-Self scale were used to define four item pairs, along with the 28 item pairs in the original seven scales described earlier. A factor analysis of responses to the 32 item pairs (see Table 3, page 42) clearly identified each of the eight factors which the revised SDQ-I was designed to measure. Again, the target loadings (in boldface) were substantial (median = .72), the nontarget loadings were small (median = .03), and the correlations among the factors were modest (median = .17). Marsh et al. employed confirmatory factor analysis to show that the first-order model described in the previous section provided an adequate fit to the data. Further, the entire solution was reasonably invariant across responses by males and females. The pattern of correlations among the original seven SDQ-I factors in Table 1 is similar to that in Table 3. The General-Self factor is modestly correlated with each of these factors but is more highly correlated with responses to nonacademic than academic factors. Nevertheless, the General-Self factor is surprisingly distinct from the other facets of self-concept, being clearly identified as a separate dimension.

At present, the General-Self scale must be interpreted cautiously. The discussion above provides a justification for its consideration, and the results of the Marsh et al. study illustrate that it can be reliably distinguished from more specific facets of self-concept. Further research presented in this Manual will examine how it is related to external validity criteria. However, neither the theoretical nor the empirical basis for General-Self scale is as well established as that of the original seven SDQ-I facets.

Table 3. Exploratory Factor Analysis Including the General-Self Scale (N = 559)

Variable	Factor							H ²	
	Physical	Appearance	Peers	Parents	Reading	Math	School		General
Factor Loadings									
Physical 1	.68	.00	.06	.04	.00	.03	.04	.00	.44
Physical 2	.51	.19	.06	.01	.01	.05	-.02	.12	.42
Physical 3	.84	.00	.09	-.01	.05	.01	-.01	.01	.61
Physical 4	.71	.02	.04	.02	-.05	-.01	.08	.13	.57
Appearance 1	-.01	.77	.03	.12	.03	.08	.02	.01	.62
Appearance 2	.02	.76	.12	.05	.03	.06	.03	.04	.66
Appearance 3	.18	.64	.11	-.05	-.01	-.06	.08	.14	.63
Appearance 4	.03	.57	.13	.00	-.03	-.04	.10	.23	.61
Peers 1	.09	.03	.72	.05	.04	.03	-.03	.03	.56
Peers 2	.03	.08	.62	.04	.02	-.01	.07	.16	.55
Peers 3	.07	.08	.65	.03	-.02	.05	.04	.08	.54
Peers 4	.05	.14	.59	.00	.02	.00	.02	.24	.62
Parents 1	.02	.04	-.02	.49	-.05	.04	.05	.03	.25
Parents 2	-.03	.10	-.03	.52	.04	.05	.09	.01	.29
Parents 3	.06	-.01	.11	.71	.01	.05	.01	.09	.52
Parents 4	.01	.01	.16	.73	-.02	-.05	.00	.10	.54
Reading 1	.02	.02	-.06	.00	.83	-.05	.07	.06	.70
Reading 2	.02	.01	.02	-.07	.85	.01	.07	.02	.72
Reading 3	.01	.01	.03	-.01	.80	.01	.07	.08	.67
Reading 4	-.05	.01	.07	.04	.77	.01	.14	.04	.67
Math 1	-.01	.02	.04	-.01	-.02	.72	.16	.08	.68
Math 2	.04	.04	-.02	.07	.02	.79	.11	.06	.73
Math 3	.02	.02	.03	.05	-.02	.76	.17	.05	.74
Math 4	.04	.01	.05	-.01	-.03	.76	.22	.03	.77
School 1	.00	.05	-.04	.04	.08	.10	.66	.11	.57
School 2	-.03	.02	-.01	-.07	.12	.19	.38	.30	.49
School 3	.03	.03	.03	.12	.06	.13	.73	-.09	.62
School 4	.06	.01	.05	-.02	.06	.14	.70	.10	.66
General 1	.10	.18	.15	.21	.06	-.03	.11	.30	.43
General 2	.03	.09	.19	.10	.01	.09	.01	.55	.55
General 3	.02	.27	.18	.02	-.03	.01	.08	.60	.60
General 4	.07	-.03	.01	.05	.07	.04	.02	.75	.53
Correlations Among Factors									
Physical	19								
Appearance	25	33							
Peers	08	18	18						
Parents	01	04	04	-.01					
Reading	07	08	07	10	.01				
Math	10	19	11	17	.25	.47			
School	23	39	43	21	16	19	.31		
General									

Note: 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. Decimal points for factor loadings and correlations are omitted.

Summary and Implications

Self-concept, like many other psychological constructs, suffers in that "everyone knows what it is." Researchers do not appear compelled to provide any theoretical definition of what they are measuring nor even the psychometric properties of the instrument used to measure it. Although many studies have been published in this area, 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 used in most self-concept research.

The multifaceted, hierarchical conceptualization of self-concept apparently is consistent with the theoretical perspectives of many researchers, and support for this perspective is apparently growing stronger. Nevertheless, the wide-spread and continued use of the "agglomerate" self-concept described earlier indicates that this approach has not been adequately recognized in empirical approaches to the measurement of self-concept. Most instruments used to measure self-concept have no clearly articulated theoretical basis, and this makes the examination of their construct validity difficult. One approach to this problem has been to take responses to existing, largely atheoretical instruments and attempt to test hypotheses from theoretical models (e.g., Marx & Winne, 1978). Because of the poor quality of measurement instruments for this purpose, the approach is dubious, and the generally inconsistent results may be due to poor theory, poor instrument construction, or both. In development of the SDQ-I, it was reasoned that the determination of whether theoretically consistent and distinguishable facets of self-concept exist, along with the nature of their content and structure, should be a prerequisite to the study of how these facets or overall self-concept are related to other variables. In adopting such an approach, atheoretical and/or purely empirical approaches to developing and refining measurement instruments were rejected. Instead, an explicit theoretical model was taken to be the starting point for instrument construction, and empirical results were used to support, refute, or revise the instrument and the theory upon which it is based. In applying this approach, the Shavelson model was judged to be the best theoretical model of self-concept. Implicit in this approach is the edict that theory building and instrument construction are inexorably intertwined and that each will suffer if the two are separated. In this sense the SDQ-I is based on a strong empirical foundation and a good theoretical model.

Reliability and stability are important psychometric properties of any measurement instrument. Even though stability is important, self-concept researchers are frequently interested in measuring change in self-concept. This chapter examines these technical issues as they pertain to the SDQ-I.

Internal Consistency Estimates

Reliability refers to the extent to which responses are due to systematic sources of variance in contrast to error variance. The primary basis for estimating reliability in SDQ research has been the internal consistency of item responses on each of the SDQ scales. Coefficient alpha estimates of reliability, based on the internal consistency of responses, have been presented in many of the published studies listed in the References. For purposes of this Manual, coefficient alphas for the SDQ-I scales and total scores were computed for the total normative sample (see Table 4). Across all the responses coefficient alphas for the eight individual scale scores varied from .80 to .92 (median = .86). Coefficient alphas for the Total Nonacademic, Total Academic, and Total Self scores were .91, .92, and .94, respectively.

Correlations among the eight positively worded items designed to measure each SDQ-I scale and corrected item-scale correlations (i.e., the correlation between an item and the sum of responses to other items in the same scale) are presented in Table 5, pages 46-48. These summary statistics further demonstrate that every item is significantly and substantially correlated with the other items designed to measure the same facet of self-concept.

Table 4. Internal Consistency Coefficients for the Total Normative Sample by Scale (N = 3,562)

Scale	Coefficient Alpha
Physical Abilities	.83
Physical Appearance	.90
Peer Relations	.85
Parent Relations	.80
Reading	.89
Mathematics	.89
General-School	.86
General-Self*	.81
Total Nonacademic	.91
Total Academic	.92
Total Self	.94

*n = 739

Analyses described in Chapter 11 indicate that self-concept varies systematically, albeit weakly, with sex and grade level. This variation is included as a source of systematic variation in the computation of reliability estimates presented in Table 4. However, reliability estimates are presented separately for males and females at different grade levels in Appendix A. Although there are some minor variations between those reliability estimates and the ones in Table 4, the differences are quite small. In summary, these results, together with those of the factor analyses summarized in Chapter 4, demonstrate that responses to the SDQ-I reliably measure facets of self-concept that are internally consistent and clearly distinguishable.

Table 5. (continued) Item Statistics by Scale (N = 3,562)

Item No.	Mean	SD	Item to Scale r	Correlations Among Items																
				2	9	16	31	39	55	63	71	General-School Item Number	General-Self* Item Number							
2	3.35	1.10	.53																	
9	3.56	1.37	.60	.34																
16	3.42	1.18	.54	.52	.32															
31	3.77	1.18	.53	.39	.32	.42														
39	3.79	1.28	.69	.35	.55	.35	.39													
55	3.04	1.32	.66	.32	.55	.31	.35	.65												
63	3.40	1.20	.89	.46	.33	.49	.50	.39	.39											
71	3.63	1.30	.71	.34	.57	.36	.36	.67	.69	.43										
29	3.74	1.14	.36																	
45	4.35	1.10	.45	.15																
53	4.22	1.07	.57	.29	.39															
67	4.23	1.05	.51	.19	.27	.33														
70	4.06	1.09	.61	.30	.33	.49	.35													
72	4.05	1.08	.61	.32	.38	.41	.35	.53												
74	4.16	1.07	.58	.21	.32	.39	.55	.41	.45											
76	4.19	0.96	.49	.25	.27	.32	.33	.39	.39	.35										

*n = 739

Stability and Systematic Changes

As mentioned in Chapter 1, self-concept researchers face an important dilemma. Theoretically, and from the perspective of measurement theory, it is desirable that self-concept be relatively stable over time, both in terms of shifts in the mean level of self-concept and in terms of stability coefficients. Some researchers also argue that it is important to a person's mental health that self-concept be stable over time. On the other hand, much of the interest in self-concept is directed toward *changes* in self-concept. Logically, dramatic life events and more gradual life changes should be reflected in changes in self-concept. Also, improvement in self-concept is often a goal of experimental interventions. Furthermore, it is often hypothesized that changes in self-concept will lead to desirable changes in other areas such as academic achievement. It is very difficult for any measure of self-concept to be perfectly stable and still be responsive to dramatic life events or systematic interventions, and herein lies the dilemma.

The stability of responses to the SDQ-I was examined by Marsh, Smith, Barnes, and Butler (1983). They examined test-retest data in two studies, one consisting of 528 fifth- and sixth-grade students, and one consisting of 143 fourth-grade students. In each study, the interval between the two testings was approximately six months, and testings took place during the same academic school year.

The stability of self-concept over time can be measured in either relative or absolute terms. *Absolute* stability implies that the average of self-concept ratings across all students does not vary over the time interval considered. There was good support for this type of stability in both studies and little change in mean scores from one test to the next. *Relative* stability can be inferred from the magnitude of correlations between self-concept measures at two different times. This type of stability is the focus of discussion here.

Change represents the other side of stability; a lack of stability suggests that some change has taken place. An important question is whether or not this change is systematic. As a reasonable approximation, the difference between the average of reliabilities determined separately at each time and the stability coefficient is an indication of systematic change; differences indicate lack of agreement between two measures beyond that which can be explained by the unreliability of measures at Time 1 and Time 2.

The results of both Marsh et al. (1983) studies are presented in the form of a multitrait-multimethod matrix where "time" is taken to be the source of method variation (see Table 6, page 51; see also Marsh, Barnes, & Hocevar, 1985; Marsh & Hocevar, 1983; Shavelson et al., 1976, for more general discussions of multitrait-multimethod analysis). Internal consistency estimates are in boldface, and the stability coefficients (or convergence coefficients in the terminology of multitrait-multimethod analyses) are in bold italics. The internal consistency of responses at Time 1 and Time 2 provides a logical upper bound for the stability over time. Internal consistency estimates were high both for the individual SDQ-I scales (mean $r = .87$) and for the total scores (mean $r = .92$). These results are similar to those shown in Table 4. Stability over time was also high for both the individual scales (mean $r = .61$) and the total scores (mean $r = .65$) with one exception, the Parent Relations scale at grade 4. There is no apparent explanation for the particularly low stability coefficient for this one scale in the fourth grade sample, though the stability of this facet was also among the lowest for the fifth- and sixth-grade samples. Although the stability coefficients were generally high, they were not nearly as high as the internal consistency estimates, which suggests that there was systematic change in student self-concepts over time.

Marsh et al. further examined the nature of the changes in self-concept responses over this six-month interval. In both studies the reliabilities of difference scores were high for both the individual scales (mean coefficient alpha = .74) and the total scores (mean coefficient alpha = .87). This finding supports the conclusion that the changes in self-concept are systematic and not due to random fluctuations. Correlations among the difference scores representing the SDQ-I scales were much smaller than their reliabilities (mean $r = .24$), and a factor analysis of the difference scores demonstrated that changes in self-concept were clearly multidimensional.

The results of the Marsh et al. study have important implications for the measurement of change in self-concept. First, self-concept was relatively stable, even for preadolescent children. Second, changes that did occur were systematic and reliable. Third, both self-concept responses and changes in self-concept over time were multidimensional and specific to particular dimensions. Thus, it appears that a particularly dramatic change in a person's life or an experimental intervention can have substantial effect on some particular component of self-concept, even if it has less effect on overall self-concept or on other specific components. However, Marsh et al. were not able to demonstrate that changes in self-concept were systematically related to changes in other variables that were considered in the study, and there is need for further research on the construct validity of changes in multiple dimensions of self-concept. In this respect the series of intervention studies by Marsh, Richards, and Barnes (1986a, 1986b) and Marsh and Richards (1988) described in Chapter 10 are important. In these intervention studies it was found that powerful interventions did produce systematic differences in self-concept. Of particular relevance to this discussion, changes in self-concept were systematically related to the goals of the program such that the scales chosen a priori as being most relevant were the ones on which the most change was observed.

Table 6. Stability and Change in SDQ-I Scale Scores over a Six-Month Interval

Scales	Grades	Intercorrelations at Time 1							Intercorrelations at Time 2						
		1	2	3	4	5	6	7	1	2	3	4	5	6	7
Time 1															
(1) Physical	5 & 6	83													
(2) Appearance	5 & 6	80	90												
(3) Peers	5 & 6	47	91	81											
(4) Parent	5 & 6	55	56	82	79										
(5) Reading	5 & 6	44	27	43	80	88									
(6) Math	5 & 6	01	13	10	29	89	90								
(7) School	5 & 6	26	20	20	24	06	92	84							
	4	31	35	35	23	19	92	84							
	5 & 6	19	27	26	35	45	55	69							
	4	33	41	50	28	43	69	82							
Time 2															
(1) Physical	5 & 6	74	31	34	01	-02	17	10	84						
(2) Appearance	5 & 6	51	32	37	06	-14	22	21	79	90					
(3) Peers	5 & 6	36	67	38	17	15	21	28	46	90	85				
(4) Parents	5 & 6	27	60	30	06	05	23	21	49	57	87	82			
(5) Reading	5 & 6	40	38	67	16	11	19	21	41	57	87	30	44		
(6) Math	5 & 6	35	49	62	26	19	26	43	41	57	87	22	44	85	
(7) School	5 & 6	10	18	17	55	15	10	12	14	22	30	44	22	15	25
	4	09	18	19	27	02	23	26	28	22	22	15	22	11	16
	5 & 6	10	06	05	19	67	-04	30	10	10	22	15	25	11	16
	4	10	23	23	21	69	07	28	11	16	30	23	19	17	91
	5 & 6	22	17	11	11	05	68	41	21	26	25	21	21	01	93
	4	28	20	24	15	07	67	45	21	26	25	21	21	01	93
	5 & 6	18	17	15	26	35	54	54	22	35	30	29	29	54	36
	4	38	34	44	23	30	59	77	39	30	48	35	40	56	88

Note: Correlations are based on unweighted sums of responses to items in each scale, with a mean test-retest interval of six months. Internal consistency estimates (coefficient alphas) are in boldface, and stability coefficients are in bold italics. Correlations of .10 and .16 are statistically significant on 5th and 6th graders ($N = 528$) and on 4th graders ($N = 143$), respectively.

Chapter 6.

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 over-all self-regard as well" (page 355). Not surprisingly, particularly in studies of school-aged children, some measure of academic achievement is one of the most frequently posited criteria used to validate self-concept interpretations, and it has also been the focus of much SDQ-I research. This chapter briefly reviews the literature, summarizes relevant research with the SDQ-I, and examines sex differences in self-concept/academic achievement relations.

In the SDQ-I and in the Shavelson model upon which it is based, self-concept is a multifaceted, hierarchically ordered construct. Academic self-concept is one component of overall self-concept, and it is broken into self-concepts in particular content areas such as math and reading. Support for the construct validity of SDQ-I interpretations and the Shavelson model requires that academic achievement be more positively correlated with academic self-concept than with nonacademic or overall self-concept, and that verbal and math achievement indicators be more highly correlated with self-concepts in matching content areas than with other facets of self-concept. In the most extensive meta-analysis of the achievement/self-concept relationship, Hansford and Hattie (1982) found that measures of ability correlated about .2 with measures of general self-concept but about .4 with measures of academic self-concept. Similarly, Shavelson and Bolus (1982) found that grades in English, mathematics, and science were more highly correlated with matching areas of self-concept than with general self-concept. Bachman (1970) reported that IQ correlated .46 with academic self-concept but only .14 with general self-concept. In her review of studies relating self-concept to academic achievement, Byrne (1984) found that nearly all studies report that self-concept is positively correlated with achievement, and many find achievement to be more strongly correlated with academic self-concept than with general self-concept. These findings support the construct interpretation described above and indicate the need to distinguish among academic, nonacademic, and general self-concepts.

Self-Concept/Achievement Correlations

SDQ research has emphasized the distinctiveness of self-concepts in verbal and mathematical content areas. For example, Marsh, Relich, and Smith (1983) demonstrated that math achievement was substantially correlated with Math self-concept (.55), less correlated with self-concepts in other academic areas (Reading, .21; General-School, .43), and nearly uncorrelated with self-concepts in nonacademic areas. The results of various studies examining the relationship between SDQ responses and academic achievement indicators are summarized in Table 7, pages 54-56, for 11 samples. Verbal, math, and general academic achievement indicators include objective tests and teacher ratings.

Table 7. Correlations Between SDQ-I Scales and Academic Performance Measures

Study ^a	Objective Test Scores			Teacher Ratings		
	Reading	Math	General Academic	Reading	Math	General Academic
Physical Abilities						
(1)	.01		-.01	-.17	-.02	-.05
(2)	-.11		-.12	-.10	.00	-.11
(3)	-.06			-.09	-.03	
(4)	-.10			-.09	-.04	
(5)	-.07			-.06	-.01	
(6)	-.03			-.05	-.04	
(7)		-.07				
(8)	-.18**	-.08		-.08	-.03	-.03
(9)	.11	-.15**				
(10)	.12	.13				
(11)	-.11**	.00				
		-.03				
Physical Appearance						
(1)	-.01		.03	-.07	-.04	-.02
(2)	-.04		.01	.02	.02	.03
(3)	-.09			-.12*	-.10	
(4)	-.10			-.07	-.01	
(5)	-.08			.01	.02	
(6)	-.03			-.01	-.07	
(7)		-.03				
(8)	-.17**	-.03		-.08	-.06	-.06
(9)	-.03	-.12*				
(10)	-.02	.08				
(11)	-.16**	-.06				
		-.05				
Peer Relations						
(1)	-.15		-.06	-.08	.03	-.02
(2)	.00		-.02	.03	.14	.05
(3)	-.08			-.08	-.07	
(4)	-.05			-.05	.00	
(5)	.03			.12	.20	
(6)	.13			.14	.14	
(7)		.09				
(8)	-.08	-.06		-.03	.03	.03
(9)	.16	.12*				
(10)	.21	.27				
(11)	-.05	.13				
		-.03				

Note: Blanks indicate that a measure was not included as part of the study.
p* < .05; *p* < .01.

Table 7. (continued) Correlations Between SDQ-I Scales and Academic Performance Measures

Study ^a	Objective Test Scores			Teacher Ratings		
	Reading	Math	General Academic	Reading	Math	General Academic
Parent Relations						
(1)	-.01		-.04	-.02	-.10	-.07
(2)	-.08		-.05	.03	.05	.01
(3)	.00			.09	.02	
(4)	.00			.01	.02	
(5)	.05			.06	.10	
(6)	.13			.08	.14	
(7)		.11				
(8)	-.11*	.03		-.05	.00	-.02
(9)	.37**	-.15**				
(10)	.07	.21				
(11)	-.11*	-.01				
Reading						
(1)	.35**		.35**	.32**	.12	.23**
(2)	.57**		.53**	.55**	.36**	.52**
(3)	.18**			.22**	.03	
(4)	.20**			.22**	.06	
(5)	.38**			.40**	.19*	
(6)	.55**			.48**	.06	
(7)		.09				
(8)	.43**	.21**		.37**	.23**	.29**
(9)	.46**	.14**				
(10)	.43**	.02				
(11)	.43**	-.01				
		.14**				
Math						
(1)	.20**		.19**	-.04	.28**	.11
(2)	.53**		.58**	.71**	.66**	.74**
(3)	.05			.13**	.28**	
(4)	.07			.09	.36**	
(5)	.00			-.04	.20**	
(6)	.04			-.01	.32**	
(7)		.34**				
(8)	-.03	.55**		.07	.28**	.21**
(9)	.03	.17**				
(10)	.03	.46**				
(11)	.03	.34**				
		.31**				

Note: Blanks indicate that a measure was not included as part of the study.
p* < .05; *p* < .01.

Table 7. (continued) Correlations Between SDQ-I Scales and Academic Performance Measures

Study ^a	Objective Test Scores			Teacher Ratings		
	Reading	Math	General Academic	Reading	Math	General Academic
General-School						
(1)	.14		.22**	.22**	.33**	.30**
(2)	.43**		.45**	.52**	.59**	.56**
(3)	.21**			.31**	.27**	
(4)	.12**			.22**	.18**	
(5)	.20*			.17*	.25**	
(6)	.26**	.23**		.22**	.29**	
(7)		.43**				
(8)	-.04	-.02		.08	.18**	.17**
(9)	.22	.46**				
(10)	.16	.21				
(11)	.14**	.26**				
General-Self						
(8)	.08	.06		.15**	.18**	.20**
(11)	.01	.08				

Note: Blanks indicate that a measure was not included as part of the study.

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

^aStudies in the Table include:

- (1) Marsh, Parker, and Smith (1983), sample 2 ($n = 180$, 6th grade);
 (2) Marsh, Parker, and Smith (1983), sample 3 ($n = 125$, 6th grade);
 (3) Marsh, Smith, Barnes, and Butler (1983), sample 1, time 1
 ($n = 528$, 5th and 6th grades);
 (4) Marsh, Smith, Barnes, and Butler (1983), sample 1, time 2
 ($n = 528$, 5th and 6th grades);
 (5) Marsh, Smith, Barnes, and Butler (1983), sample 2, time 1
 ($n = 143$, 4th grade);
 (6) Marsh, Smith, Barnes, and Butler (1983), sample 2, time 2
 ($n = 143$, 4th grade);
 (7) Marsh, Reilich, and Smith (1983), ($N = 498$, 6th grade);
 (8) Marsh, Smith, and Barnes (1984), ($N = 559$, 5th grade);
 (9) Marsh and Richards (1986), Time 1 ($N = 43$, 9th grade);
 (10) Marsh and Richards (1986), Time 2 ($N = 43$, 9th grade);
 (11) Marsh and Gouvernet (in press), ($N = 508$, 7th, 8th, and 9th grades).

Table 7 presents 144 correlations between academic achievement indicators and the four nonacademic SDQ facets. Few of these correlations reach statistical significance, most are negative, and only one correlation is significantly positive. Given the range of studies and the diversity of indicators of academic achievement, these results provide convincing evidence for the relative independence of academic achievement and the nonacademic self-concept scales.

Tests of verbal achievement and teacher ratings of reading achievement have been the most frequently employed achievement indicators in SDQ research. The 17 correlations between Reading self-concept and verbal achievement indicators vary from .18 to .57 (median = .40), and all are statistically significant. The median correlation between these same verbal achievement measures and Math self-concept is .04, and only 4 of 17 correlations are statistically significant. The 17 correlations between reading achievement indicators and General-School self-concept vary from -.04 to .52 (median = .21), 12 of which are statistically significant. In summary, with the possible exception of the anomalous reading achievement/Math self-concept correlation in Study 1 (Sample 3), these results indicate that reading achievement indicators are most highly correlated with Reading self-concept, less correlated with the General-School scale, even less correlated with Math self-concept, and uncorrelated or negatively correlated with self-concepts in nonacademic areas.

Math achievement indicators were collected in fewer studies, but the 13 correlations between math achievement and Math self-concept vary from .17 to .66 (median = .32), all statistically significant. The 13 correlations between math achievement and Reading self-concept vary from -.01 to .36 (median = .12), and six are statistically significant. The 13 correlations between math achievement and General-School self-concept vary between -.02 and .59 (median = .26), and 11 are statistically significant. Thus, math achievement indicators are most highly correlated with Math self-concept, less highly correlated with General-School self-concept, even less correlated with Reading self-concept, and uncorrelated or negatively correlated with self-concepts in nonacademic areas.

In summary, the correlations between SDQ scales and academic achievement indicators support a dramatic distinction between academic and nonacademic facets of self-concept and also demonstrate the clear separation of Math and Reading self-concepts.

Sex Differences in Self-Concept/Achievement Relations

Sex differences in achievement and self-concept have been extensively examined only in the area of mathematics. In a literature review Meece, Parsons, Kaczala, Goff, and Futterman (1982) reported that sex differences in math achievement and math self-concept are not large in the elementary school years. Females generally do as well as males on standardized tests of math achievement during elementary and junior high school (e.g., Fennema, 1974; Sherman, 1980). However, some studies indicate that by junior (early) and senior (late) high school years, females have lower levels of math achievement and self-concept (Meece et al., 1982). The Meece et al. review suggests that as students go through high school, math self-concept declines, but the decline begins sooner and is larger for females than for males. They assert that the decline in females' math self-concepts precedes the decline in their math achievement and that socialization processes reflected in math self-concepts are one cause of the decline in achievement. Particularly persuasive support for this assertion would be a demonstration that females have significantly lower math self-concepts even though their actual school performance and/or achievement in

math equals or surpasses that of males. Extrapolating from the Meece et al. review, such an occurrence would be most likely during late primary or early junior high school years, and the examination of such a possibility is the purpose of discussion in this section.

The results described here are based on a study by Marsh, Smith, and Barnes (1985). In this study, 559 fifth-grade students completed the SDQ-I and achievement tests in math and reading administered by the researchers. They were also evaluated in terms of their self-concepts and achievement in different areas by their classroom teachers. The study found that females had significantly lower self-concepts than males in Physical Abilities, Physical Appearance, and Math, significantly higher self-concepts in Reading, and did not differ significantly from males (at $p < .01$) for Peer Relations, Parent Relations, General-School, and General-Self. Because of the large sample size, even small correlations were statistically significant, but sex differences accounted for no more than 3% of the variance in any of the self-concept scores. In addition, for both test scores and teacher ratings females had higher achievement scores in math and reading. These findings indicate that these females in late primary grades had significantly higher levels of mathematics achievement but significantly lower levels of Math self-concept.

The relationships among sex, reading and math achievement, and Reading and Math self-concepts were further examined in a series of path analyses (see Figure 12). In Models 1 through 3, achievement test scores, teacher ratings of student achievement, and their totals are related to Reading and Math self-concepts. In Models 4 and 5, sex is added to variables considered in the first three models. For reading, females had higher levels of achievement, leading to better Reading self-concept. Sex had no direct effect on Reading self-concept; the higher Reading self-concept for females could be explained in terms of their higher levels of achievement (e.g. see Model 4 in Figure 12). However, the pattern was quite different for mathematics. Females had significantly higher levels of math achievement but lower levels of Math self-concept. Thus, being female had a positive direct effect on math achievement but a negative direct effect on Math self-concept. These results were based on academic achievement assessed by test scores, but the same pattern of results was also observed for achievement based on teacher ratings and on the total of teacher ratings and academic achievement.

In Model 5 (Figure 12) teacher ratings and test scores were considered separately, and it was assumed that academic achievement as reflected in objective tests is one of the causal determinants of teacher ratings of academic ability. According to this model, being female had a direct positive effect on both reading test scores and teacher ratings of reading. Teachers rated females' reading ability to be higher, and this sex difference was larger than could be accounted for by differences in reading test scores. Being female also had a direct positive effect on math test scores but not on teacher ratings of math; the higher teacher ratings of females' math ability could be accounted for by differences in test scores. Similarly, being a female had no direct effect on Reading self-concept but had a negative direct effect on Math self-concept. The effects of the teacher ratings are important since teachers perceived the females to be academically more able at math, and thus it is unlikely that the lower Math self-concepts for the females were due to teacher expectations.

The support for the Meece et al. proposal demonstrated here is particularly compelling since females were shown to have significantly lower self-concepts in mathematics even though their level of academic achievement was higher as measured by objective test scores and as judged by their teachers. Because the direction of the sex effects on the two variables was opposite, there is no need to demonstrate the causal predominance of one effect over the other, and many alternative explanations (e.g., that changes in self-concept are produced by changes in achievement) are not viable.

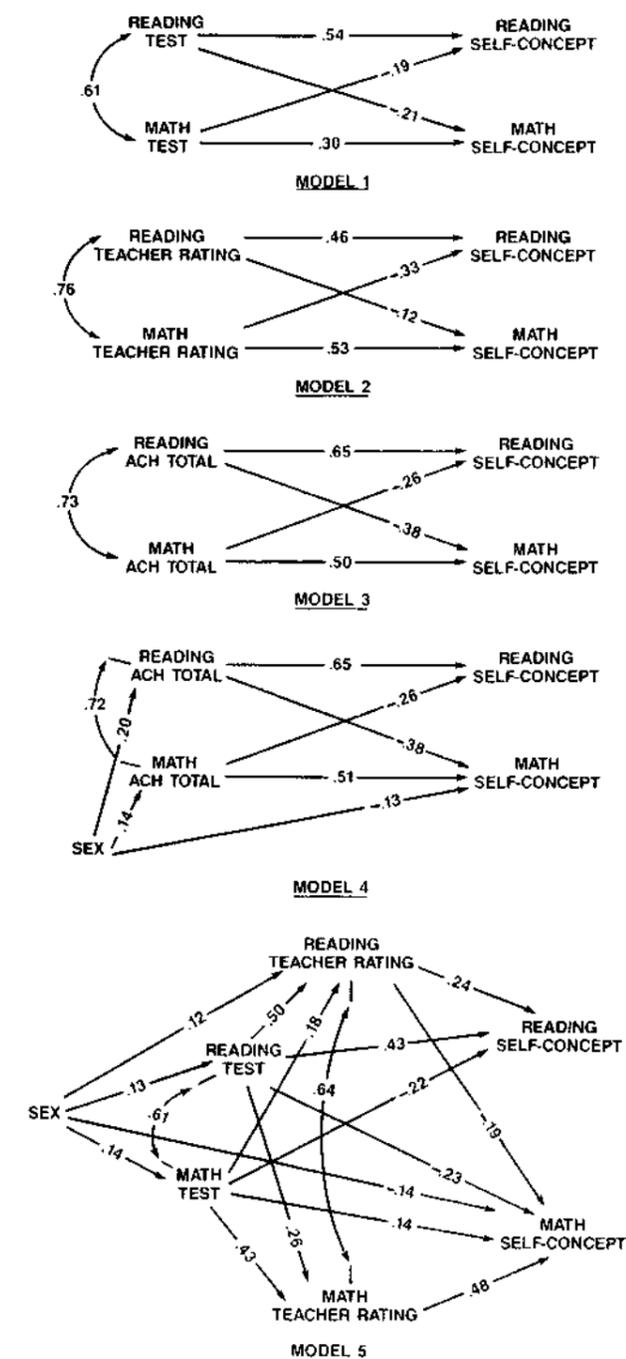


Figure 12. Path Models of Relationships Among Gender, Verbal and Math Achievement Indicators, and Reading and Math Self-Concepts

From Marsh, H.W., Smith, I.D., & Barnes, J. (1985). Multidimensional self-concepts: Relations with Sex and Academic Achievement. *Journal of Educational Psychology*, 77, p. 590. Copyright 1985 by the American Psychological Association. Reprinted by permission of the publisher.

In another Australian study with the SDQ-I in schools similar to those considered here, Relich (1983) also found that sixth-grade females had significantly lower Math self-concepts than did males even though the females had significantly higher levels of mathematics achievement. Although the focus of the Relich dissertation was not on self-concept, his findings provide corroboration for the sex differences observed here. These findings, in conjunction with the Meece et al. review, are important in that they provide a compelling demonstration that changes in self-concept apparently can lead to (i.e., cause) subsequent changes in academic achievement. Nevertheless, support for causal interpretations of correlational results must be viewed cautiously. For example, the females in this study were not actually shown to have subsequently lower math achievement scores in high school. Thus, there is a need for further examination of this conclusion in longitudinal studies.

Sex differences observed here were generally consistent with findings from other studies and supported observations from Meece et al. in their review of mathematics achievement and affect. Nevertheless, they cited no research in which, like the Marsh, Smith, and Barnes (1985) study, females had higher achievement scores in math and reading and higher self-concepts in reading but lower self-concepts in math. They did argue that sex differences favoring males in mathematics self-concept preceded sex differences in mathematics achievement and that sex differences favoring males in mathematics achievement are not well established until junior and senior high school. This observation is important in their schema, since it was one basis for their contention that sex differences in mathematics achievement are due to stereotyped socialization patterns which produce traditional sex roles, attitudes, and beliefs. We interpret this to mean that socialization as reflected in self-concept differences produces achievement differences.

Marsh, Byrne, and Shavelson (in press) found that 11th- and 12th-grade males had significantly higher Math self-concepts as measured by the SDQ-III than did 11th- and 12th-grade females, while females had higher Verbal self-concepts than males. These sex differences, though weaker, were still evident even after controlling for school grades in English and mathematics. As reported in other studies, being female contributed to a lower Math self-concept than could be explained by school performance in mathematics. However, being male contributed to a lower Verbal self-concept than could be explained by school performance in verbal areas. Although the pattern of relations between females and mathematics has been well publicized, fewer studies have examined what may be a corresponding pattern of relations between males and verbal skills.

Chapter 7.

Frame of Reference Effects on Self-Concept and Achievement

A considerable body of research indicates that academic self-concepts are at least moderately correlated with corresponding levels of academic achievement (see Chapter 6). However, the correlations almost never approach the reliabilities of the respective measures, suggesting that academic self-concepts reflect more than just academic achievement. Even after individuals obtain information from various sources about their levels of academic ability or achievement, these impressions must be compared to some standard or frame of reference. To the extent that individuals have different frames of reference, the same objective indicators of academic ability will lead to different academic self-concepts.

Theoretical models and empirical support for two different frame of reference effects are examined in this chapter. In the first it is hypothesized that academic self-concepts are substantially influenced by the ability levels of other students in the immediate context as well as by one's own ability level. Thus, average-ability students will have higher academic self-concepts in low-ability schools than in high-ability schools. This is called the "Big Fish Little Pond Effect." In the second it is proposed that in order to formulate their academic self-concepts, students compare their own ability levels in different academic subjects in addition to comparing their ability levels to those of other students. Thus, a student who is poor in all academic areas but relatively better in mathematics than in other school subjects may have an average Math self-concept.

The Big Fish Little Pond Effect

Marsh and Parker (1984; also see Marsh, 1984b, 1984c) originally sought to replicate studies in the United States which found that students in low-socioeconomic status (SES)/low-ability schools have higher self-concepts than do students in high-SES/high-ability schools (Soares & Soares, 1969; Trowbridge, 1970, 1972). Researchers sometimes assume that disadvantaged children are likely to have lower self-concepts, and these earlier studies were particularly important because they countered this assumption.

Wylie (1979), in the most extensive review of research in this area, reported that "48 studies involving both well-known and idiosyncratic instruments to index overall self-regard have yielded contradictory, weak, mostly null results regarding the relationship of socioeconomic status and overall self-regard" (page 115). In the two frequently cited studies by Soares and Soares (1969) and Trowbridge (1972), Wylie found what she called a paradoxical relationship in which SES was negatively correlated with overall self-concept. These two studies, unlike most, were based on school-average SES rather than individual SES. Wylie noted that there was better evidence for a positive correlation between *individual* SES and *academic* self-concept. Brookover and Passalacqua (1981) also reported that, although individual

academic achievement is positively correlated with individual measures of academic self-concept, school-average measures of academic achievement were negatively correlated with self-concept. Thus, it appears that the size and perhaps even the direction of correlation between self-concept and academic achievement may depend on the particular facet of self-concept and on the level of data aggregation (i.e., individual vs. school-average).

Two large longitudinal studies (Bachman & O'Malley, 1977; Maruyama, Rubin, & Kingsbury, 1981) examined the relationships among academic achievement, SES, and self-concept. The researchers proposed the causal ordering among these variables to be: (a) family SES, (b) academic ability based on objective test scores, (c) school grades, (d) overall self-concept, and (e) subsequent criterion measures. Both of these studies reported only modest effects of SES and academic achievement indicators on overall self-concept. However, both studies failed to consider two key issues — neither distinguished between academic and nonacademic self-concepts, and neither considered school-average variables representing SES or academic ability (see Marsh & Parker, 1984, for further discussion). The Bachman and O'Malley study was particularly notable because measures were systematically collected from a large number of different schools so that school-average measures could be computed, and academic self-concept was assessed so that it was available to be added to their analyses.

Description of the Model

The Big Fish Little Pond Effect (BFLPE) occurs when equally able students have lower academic self-concepts when they compare themselves to more able students and higher academic self-concepts when they compare themselves with less able students (Marsh, 1984b, 1984c, 1987a; Marsh & Parker, 1984). The frame of reference model, designed to explain the BFLPE, hypothesizes that students compare their own academic ability, more or less accurately perceived, with their perceptions of the academic ability of other students in their reference group. Students then use this relativistic impression as one basis for forming their own academic self-concept.

A possible representation of this model is presented in Figure 13 for students X, Y, and Z who differ in terms of their objective academic ability. Based on the total sample, student Y has an average level of academic ability. However, if Y attends a high-ability school, this level of academic ability will be below the average of other students in the school, leading to an academic self-concept that is below average. On the other hand, if Y attends a low-ability school, the same level of academic ability will be above the average in this school and will lead to an academic self-concept that is above average. Similarly, the academic self-concepts of students X and Z will depend on their objective academic ability but will also vary with the type of school they attend. According to this model, academic self-concept will be positively correlated with individual achievement and variables related to it (e.g., family SES) but *negatively* correlated with school-average achievement and variables related to it (e.g., school-average SES).

Support for the Model

Marsh and Parker (1984) sampled sixth-grade classes selected from high-SES and low-SES schools within the same city. The two samples differed substantially in terms of average family property values, average family income levels, job occupational status of the family's principal wage earner, reading achievement scores, and

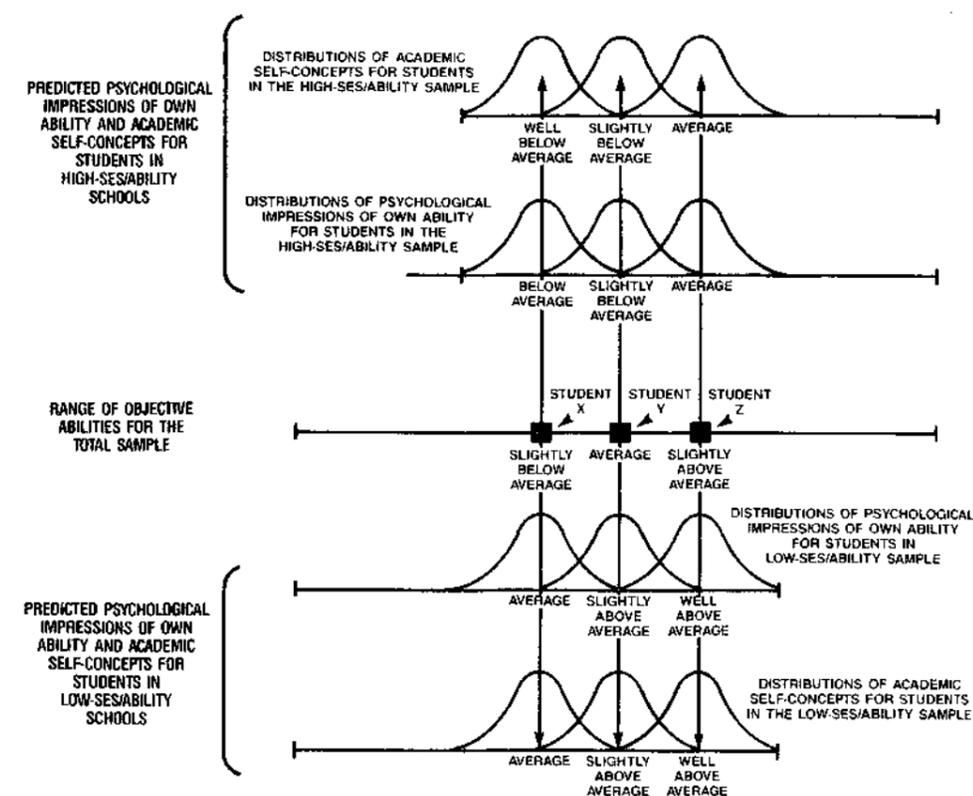


Figure 13. Theoretical Model of the External Frame of Reference Hypothesis

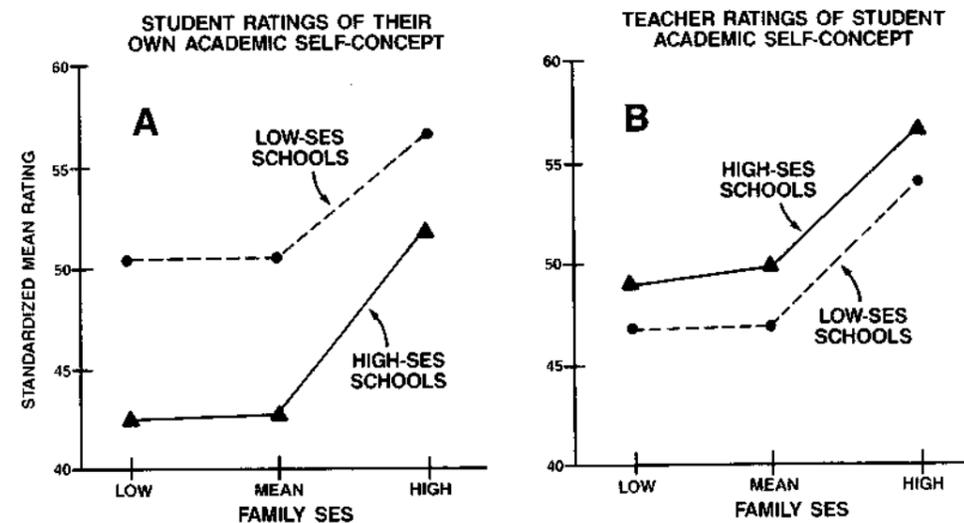
From Marsh, H.W. (1984). Self-concept: The application of a frame of reference model to explain paradoxical results. *Australian Journal of Psychology*, 28, p. 178. Copyright 1984 by H.W. Marsh. Reprinted by permission of the author.

IQ scores. Because of the way in which the group was sampled, school-average ability and school-average SES were almost perfectly correlated. For this reason the Marsh and Parker study was not able to differentiate between the effects of these two school-average variables. Thus, the effect is referred to as school-average ability/SES in subsequent discussion of this study.

For purposes of the study, students completed the SDQ-I and a reading achievement test, teachers rated the students in terms of academic abilities and inferred self-concepts, and IQ scores were made available from school records. In the first analysis, two-way ANOVAs were conducted to test the relationships among family SES (categorized into three levels), school-average ability/SES, and self-concept scores. Neither family SES nor school-average ability/SES were significantly related to nonacademic self-concept, but both variables were significantly related to academic self-concept (see Figure 14, page 64). Paradoxically but consistent with the BFLPE, the direction of the effects of family SES and school-average ability/SES on academic self-concept was opposite. Within each level of school-average ability/SES, the higher the family SES, the higher the academic self-concept (i.e., family SES, corrected for school-average ability/SES, was positively related to academic self-concept). In contrast, at each level of family SES, higher academic self-concepts were found in the low-ability/SES schools (i.e., school-average ability/SES, corrected for family SES, was negatively related to academic self-concept).

Despite the fact that teacher ratings of student (inferred) self-concepts were substantially correlated with student self-concepts, the effects of the SES variables on teacher ratings were quite different. For self-concepts inferred by teachers, higher levels of family SES and school-average ability/SES were associated with higher academic and nonacademic self-concepts. Thus, teachers in high-ability/SES schools judged the self-concepts of their students to be higher, whereas for students in a high-ability/SES school, lower academic self-concepts resulted. This suggests that teachers and students may be using different frames of reference as a basis of their judgments. As predicted by the frame of reference model, students seemed to be comparing themselves with other students in their own school, whereas teachers were using a broader, more absolute frame of reference. As a result students with average academic abilities see themselves as "below average" in high-ability/SES schools and "above average" in low-ability/SES schools, and they use this relativistic impression as one basis of their academic self-concept. Teachers, on the other hand, judge students to have lower academic self-concepts in low-ability/SES schools and higher academic self-concepts in high-ability/SES schools. Consistent with other SDQ-I research, it also appears that teachers less clearly differentiate between academic and nonacademic self-concepts than do students.

A series of path models was developed to further explore these paradoxical findings (see Figure 15, page 66). Model I shows the relationship between student academic ability, school-average ability, and the two self-concepts. Here the direct effect of academic ability on academic self-concept was positive, but the direct effect of



The relations between the two SES variables and academic self-concept as judged by students themselves (Panel A) and by their teachers (Panel B). [$M = 50$, $SD = 10$ for both sets of ratings. The sample sizes differ because a few students—14 of 305—were not judged by their teachers. For Panel A, school SES, $F(1, 299) = 11.9$, $p < .01$; family SES, $F(2, 299) = 8.7$, $p < .01$; interaction effect, $F(2, 299) = 1.8$, $p > .2$. For Panel B, school SES, $F(1, 286) = 7.0$, $p < .01$; family SES, $F(2, 286) = 9.6$, $p < .01$; interaction effect, $F(2, 286) = 0.7$, $p > .2$.]

Figure 14. Relationships between SES and Academic Self-Concepts and Inferred Academic Self-Concepts

From 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, p. 223. Copyright 1984 by the American Psychological Association. Reprinted by permission of the publisher.

school-average academic ability was negative. In Model II the relationships among family SES, school SES, and the two academic measures show a similar pattern in which the direct effect of family SES on academic self-concept was positive while the effect of school SES was negative. In Model III the effects of both SES variables and student academic ability are examined simultaneously. In this model the direct effect of school SES was even more negative than in Model II. This finding is important because for a given student both family SES and academic ability are relatively constant and thereby "controlled," which is important in assessing the negative effect of school SES on academic self-concept. It is also interesting to note that school SES positively affected academic achievement even though it was negatively related to academic self-concept. This suggests that, for a given student, being in a low-ability school produces a higher level of academic self-concept even though it may result in a somewhat lower level of academic performance.

The Marsh and Parker (1984) study was designed to replicate the controversial findings by Soares and Soares (1969) and Trowbridge (1972) in which students in low-SES schools had higher self-concepts than students in high-SES schools. In all three studies, the negative effect of school SES on total self-concept was statistically significant but very small (r 's between $-.07$ and $-.13$). However, in contrast to these weak zero-order correlations, the negative effect of school SES on academic self-concept is much stronger when the effects of family SES and student academic ability are controlled ($-.36$ for controlled vs. $-.08$ for uncontrolled). Thus, not only do the findings of this study replicate the earlier ones, but also a more detailed analysis demonstrates that the negative impact of school SES was seriously underestimated by the earlier studies.

Additional studies of the effect of school-average academic ability and school-average SES lend further support to the model. Marsh and Parker (1984) proposed a reanalysis of the Youth in Transition data from the Bachman and O'Malley (1977) study that included Bachman and O'Malley's measure of academic self-concept and school-average measures of academic ability and SES. The Youth in Transition data comprise one age cohort of 2,213 males selected on the basis of a multistage probability sample of 87 schools representing all public schools in the 48 states of the continental United States. The subjects were tested on four occasions from grade 10 to about 5 years after graduation from high school. Thus, the data are well suited for testing the generalizability of the BFLPE in a broader, more representative sample of older students. Bachman and O'Malley (1986) conducted such a reanalysis from which all nonwhite students and predominantly nonwhite high schools were excluded. Their results provided support for both the BFLPE and the construct validity of academic self-concept. However, the size of the BFLPE in their study was much smaller than reported by Marsh and Parker (1984).

The frame of reference model used to explain the BFLPE predicts that the size of the effect will vary according to the variability of school-average ability. Marsh and Parker selected schools that appeared to be extreme in terms of school-average ability. This might have increased the size of the BFLPE in their study relative to the estimate that would have been found with a more representative sample of schools. In contrast Bachman and O'Malley (1986) excluded all nonwhite students and all predominantly nonwhite schools. Thus, the variability of school-average ability in their sample was substantially smaller than in their total sample. The model predicts that the size of the BFLPE should be substantially smaller in the less variable subsample than in the total sample. Marsh (1987a), in reanalysis of the Youth in Transition total sample and subsample, provided empirical support for this prediction. It is emphasized that these differences are explicable in terms of the same theoretical model that is used to explain the BFLPE and thus provide support for the model. These complications make it difficult to establish the size of the BFLPE in an absolute sense, but they do illustrate that its size will depend on the particular application.

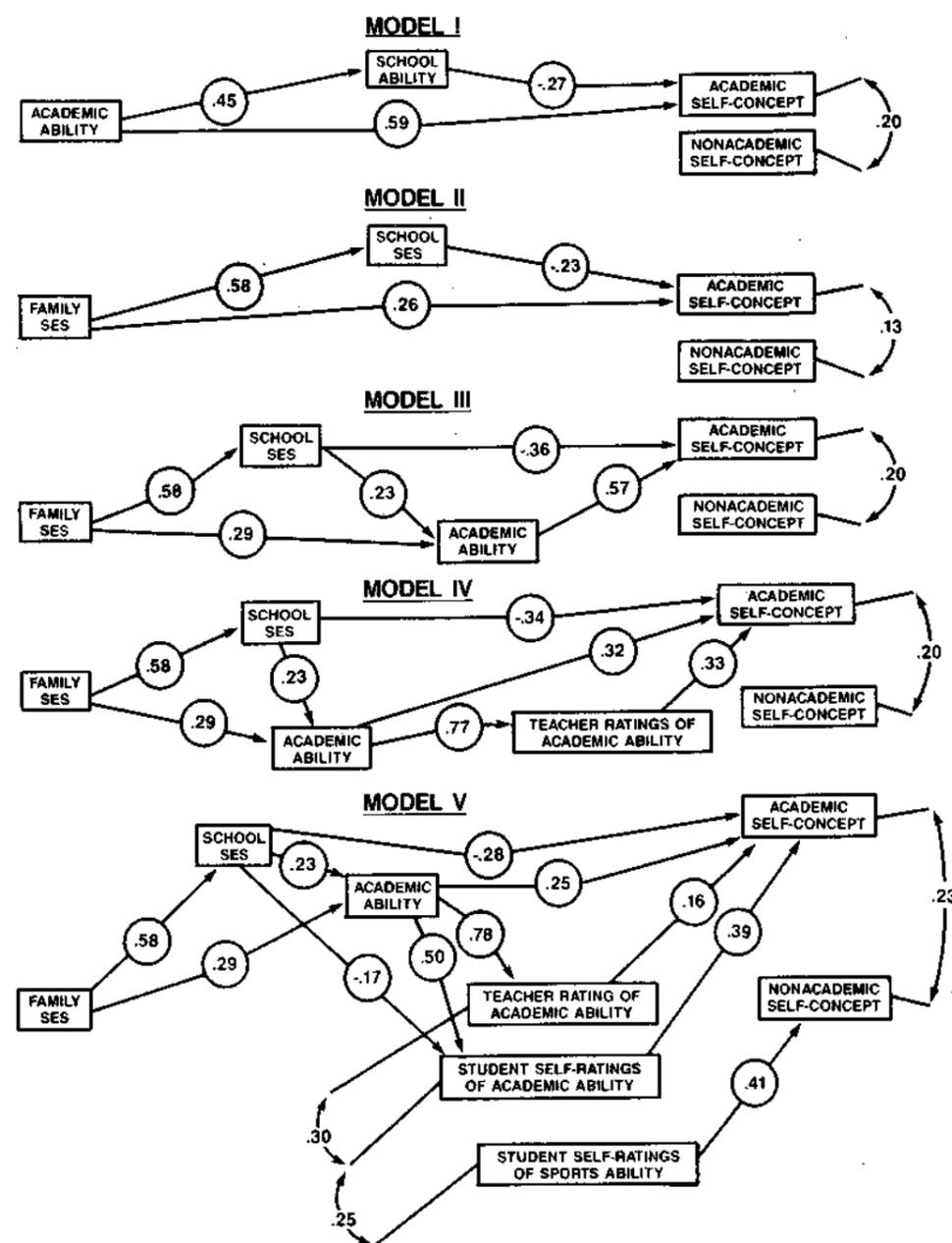


Figure 15. Path Models of Relationships Among SES, Academic Ability, and Student Self-Concepts

From Marsh, H.W. (1984). Self-concept: The application of a frame of reference model to explain paradoxical results. *Australian Journal of Psychology*, 28, p. 173. Copyright 1984 by H. W. Marsh. Reprinted by permission of the author.

Black students in the Youth in Transition study, particularly those in segregated schools, had substantially lower academic ability scores than white students. However, their academic self-concept scores were only slightly lower than those of whites. Though this pattern might suggest that responses are biased, this is exactly the pattern of results predicted to occur in the BFLPE. Blacks had academic ability test scores that were below average; however, particularly in the segregated schools, they compared themselves to classmates who also had below-average test scores. Thus, although their academic self-concepts were somewhat below average (due perhaps to self-perceptions that were independent of the immediate school context), they were not nearly as low as the ability tests would have predicted. This explanation of black-white differences in terms of the BFLPE offers empirical support for a theoretical explanation which does not assume that responses by blacks or whites are biased or differentially affected by response styles.

The results of Marsh's (1987a) reanalysis provide further insight into the distinction between academic ability and grade point average (GPA), their respective influences on academic self-concept, and how this is influenced by frame of reference effects. The 87 schools in the study differed substantially in terms of school-average academic ability but not school-average GPA. Apparently schools grade on a curve, such that the distribution of grades is similar from one school to the next even when actual academic ability levels are not (also see Davis, 1966). Thus, a substantial frame of reference effect influences GPA independent of academic ability: equally able students have lower GPAs in high-ability schools than in low-ability schools. This frame of reference effect is separate from, but contributes to, the BFLPE on academic self-concept. The fact that ability test scores contributed strongly to academic self-concept in addition to their indirect effect through GPA suggests that students do have a substantial basis beyond the information provided by GPA for inferring their academic ability levels. These external sources of inference do contribute to academic self-concepts.

The Marsh (1987a) study also clarified the effects of school-average SES and school-average ability. The theoretical basis of the BFLPE suggests that the effect should be based on school-average ability rather than school-average SES. Consistent with this prediction, when both school-average measures were included in the same model, the effects of school-average ability on academic self-concept were consistently negative, whereas the effects of school-average SES were negligible.

Marsh (1987a) offered several hypotheses about the BFLPE on the basis of his study. First, the BFLPE is primarily a function of school-average ability rather than school-average SES. Second, the size of the BFLPE will vary according to the variability of school-average ability in the particular sample. Third, although there is no clear empirical support, the size of the BFLPE may be smaller for older students. Fourth, differences between black and white students in academic self-concept are primarily due to the BFLPE and not due to racially determined response biases. Finally, a similar sort of frame of reference effect inherent in GPA affects the grades that are assigned to students in schools of differing academic abilities (unless grades are externally moderated). This effect contributes to the BFLPE.

The existence of the BFLPE is supported by other research using a variety of experimental/analytical approaches. Rogers, Smith, and Coleman (1978) ranked a group of children in terms of academic achievement across the whole group and then within their own classrooms (i.e., relative to their classmates rather than to the larger, more representative sample). They found that the within-classroom rankings were more highly correlated with self-concept. Strang, Smith, and Rogers (1978) tested the self-concepts of academically disadvantaged children who attended some classes with other disadvantaged children and some classes with nondisadvantaged children. These academically disadvantaged children were randomly

assigned to experimental and control groups. The experimental group was manipulated to enhance the saliency of their membership in the regular classrooms with nondisadvantaged children, and these children reported lower self-concepts than the control group.

Schwarzer, Jerusalem, and Lange (1983; also see Jerusalem, 1984) examined the self-concepts of West German students who moved from nonselective, heterogeneous primary schools to secondary schools that were streamed on the basis of academic achievement. At the transition point students chosen to enter the high-ability schools had substantially higher academic self-concepts than those entering the low-ability schools, but the two groups did not differ in academic self-concept by the end of their first year in the new schools. Path analyses indicated that the direct influence of school type on academic self-concept was negative. In a meta-analysis of studies of the effect of homogeneous ability grouping on self-concept, Kulik (1985; also see Kulik & Kulik, 1982; Marsh, 1984c) found that high-ability students tended to have lower self-concepts and low-ability students tended to have higher self-concepts, when placed in streamed classes of students with similar abilities than did unstreamed comparison groups. In summary, each of these studies provides support for the BFLPE in that one's own academic self-concept is negatively related to the average performance of classmates.

Davis (1966) posited a model similar to the BFLPE model in a study of the career decisions of college males. He sought support for a theoretical explanation of why the academic quality of a college had so little effect on career choice. He proposed that attending a high-ability college would result in a poorer GPA independent of individual academic ability and that GPA influences self-evaluations and, subsequently, career decisions. Davis could not fully test his model because he had no individual measures of academic ability, but he did receive dichotomous responses to the item "I have a flair for course work in this area" as an indicator of self-evaluations in eight subject areas. Also, using a self-reported measure of GPA, Davis found GPA to be more strongly related to both career decisions and "flair" than was school-average ability, and that "flair" contributed uniquely to career decisions. Furthermore, "flair" in particular subject areas was logically related to particular career decisions. For example, choosing the physical sciences as a career was positively related to "flair" in physical science and mathematics courses but negatively related to "flair" in biological sciences, social sciences, and English after controlling for GPA. Davis concluded: "The aphorism, 'It is better to be a big frog in a small pond than a small frog in a big pond' is not perfect advice, but it is not trivial" (page 31).

Discussion

In addition to having important theoretical implications for understanding self-concept and its measurement, the BFLPE has important practical implications. For instance, consider the question of whether it is better to send a child to a school where the average ability of other students will be high, or to a low-ability school. In terms of academic self-concept, it is better to be in a low-ability school, since this will lead to a higher level of academic self-concept. At least for some children, the early formation of self-image as a poor student may be more detrimental than the possible benefits of attending a high-ability school. This creates a dilemma for parents and is becoming more frequently encountered as dissatisfaction with public schools becomes stronger. Particularly for middle- and upper-middle-class families who live in inner cities, many parents must decide whether to send their children to local schools where school-average ability may be low, or to selective, high-ability schools. It is also important to note that the size of the BFLPE based on a representative sample underestimates the size of the BFLPE relative to this particular decision when school-average ability differences are large. Similarly,

Davis (1966) warned that "Counselors and parents might well consider the drawbacks of sending their boy to a 'fine' college, if, when doing so, it is fairly certain that he will end up in the bottom ranks of his graduating class" (page 31).

Marsh and Parker (1984) cautioned that a positive self-concept which is based on comparisons with the abilities of others in a low-ability school may not be maintainable in a different academic setting. Marsh (1984a) later described a dynamic equilibrium model in which academic achievement, academic self-concept, and attributions for the causes of academic success and failure are interwoven in a network of reciprocal relations, to the extent that a change in any one will produce changes in the others in order to reestablish an equilibrium (see Chapter 8). For example, some students moving from a low-ability school to a high-ability school might lower their academic self-concepts, some might improve their academic performance, some might change their academic attributions so as to protect their previous academic self-concepts, and some might use various combinations of these possibilities. There is a need for research into what actually happens when students move from one academic setting to another where the average ability level is quite different and also into the individual characteristics that may determine how students react to this potentially stressful transition.

BFLPE research reviewed in this chapter indicates that school-average ability negatively affects academic self-concept. Implications from this research assume that, at least for some children, these negative effects on academic self-concept will outweigh potential benefits from attending a high-ability school. Implicit in these assumptions is the yet untested assumption that there are *benefits* from attending a high-ability school. Marsh (1988a) more fully explored this assumption in a longitudinal investigation of a large, nationally representative sample of students in the High School and Beyond study. He examined a variety of academic outcomes (e.g., standardized examination performance, academic self-concept, selection of advanced course work, time spent on homework, quality of academic effort, school grades, postsecondary attendance) measured in the sophomore and senior years of high school and two years after the normal graduation from high school. Using a path analytic strategy, he found that the effect of school-average ability on academic self-concept was negative, thus adding evidence for the generalizability of the BFLPE. He also found that the influence of school-average ability was not positive for any of the 14 outcome variables that were considered, and it was moderately negative for some (the negative effect on academic self-concept was the largest effect). He concluded that the academic outcomes produced by attending high-ability schools were not even commensurate with the initial high ability levels of students who attended those schools and that no academic advantages of such schools were observed for the variables that were considered. It must be emphasized that the size of the negative effects of school-average ability were typically small and that the results were averaged over 1,000 high schools and many thousands of students. Thus, there will be some high-ability schools that produce academic outcomes consistent with the quality of their high-ability students and some students who will be advantaged by attending such high-ability schools. Nevertheless, the findings do demonstrate that it is unjustified to assume that attending high-ability schools will necessarily result in any academic advantages.

Internal/External Frame of Reference Model

SDQ-I research based on responses by children and preadolescents, as well as SDQ-II and SDQ-III research based on responses by adolescents and young adults, have consistently found little correlation between Reading and Math self-concepts. The purpose of the present discussion is to examine empirical support for the internal/external frame of reference model (the I/E model). This model describes

relations between Reading and Math self-concepts and between these academic self-concepts and verbal and math achievement. See Marsh (1986d) for a more complete discussion of this research.

Achievement/ability measures in verbal and mathematical areas typically correlate from .5 to .8, so it is reasonable to expect that the self-concepts will also be substantially correlated. This expectation was incorporated into the original Shavelson model, in which academic self-concepts in particular subject areas were posited to form a general academic self-concept. Thus, it is surprising that Math and Reading self-concepts have been found to be nearly uncorrelated with each other. This unexpected lack of correlation has been observed in several studies with various SDQ instruments, and Marsh (1986d) proposed a theoretical model to explain its occurrence. This finding also led to a revision of the Shavelson model (Marsh & Shavelson, 1985; Shavelson & Marsh, 1986) in which self-concepts in particular subject areas are posited to form verbal/academic and mathematical/academic self-concepts (see Chapter 4).

According to the I/E model, Reading and Math self-concepts are formed in relation to both external and internal comparisons or frames of reference:

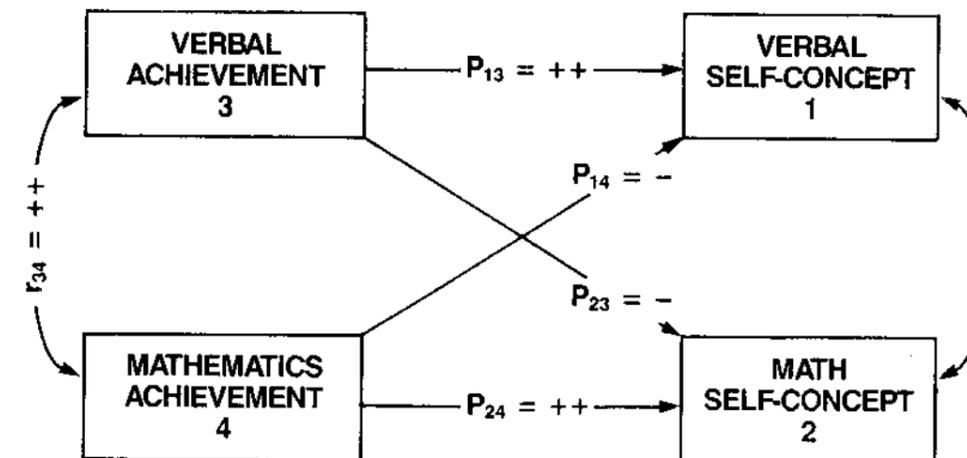
- **External Comparisons:** According to this process, students compare their perceptions of their own math and reading ability with the perceived abilities of other students within their frame of reference. They use this external relativistic impression as one basis of their self-concept in each of the two areas. (This is the same process described earlier to explain the BFLPE.) It is also assumed that this process is used by external observers to infer the self-concept of someone else.
- **Internal Comparisons:** According to this process, students compare their self-perceived ability in math with their self-perceived ability in reading and use this internal, relativistic impression as a second basis for their self-concept in each of the two areas.

To clarify how these two processes operate, consider a student who accurately perceives himself or herself to be below average in both math and reading skills but who is better at math than at reading and other academic subjects. This student's math skills are below average relative to other students' (an external comparison) but *higher* than average relative to his or her skills in other academic areas (an internal comparison). Depending upon how these two components are weighted, this student may have an average or even above-average self-concept in mathematics despite his or her poor math skills.

The external process was supported in BFLPE studies. Because reading and math achievements are substantially correlated, this external comparison process should lead to a positive correlation between Reading and Math self-concepts. However, the internal process should lead to a negative correlation between Reading and Math self-concepts, since math and reading abilities are compared with each other, and it is the difference between math and verbal skills that contributes to a higher self-concept in one area than in the other. The external process predicts a positive correlation between Reading and Math self-concepts, the internal process predicts a negative correlation, and the joint operation of both processes, depending upon the relative strength of each, will lead to the near-zero correlation between Reading and Math self-concepts that has been observed in empirical research. The I/E model does not require that the Reading/Math self-concept correlation be zero, but it does require that it be substantially less than the typically large correlation between verbal and math achievement levels.

The I/E model also predicts a *negative* direct effect of mathematics achievement on Reading self-concept and of reading achievement on Math self-concept. For example, a high Math self-concept will be more likely when math skills are good (the external comparison) *and* when math skills are better than reading skills (the internal comparison). Thus, once math skills are controlled, it is the *difference* between math and reading skills which is predictive of Math self-concept, and high reading skills will actually detract from a high Math self-concept.

The I/E model generates a specific and perhaps unexpected pattern of relations among variables representing Reading self-concept, Math self-concept, verbal achievement, and math achievement (see Figure 16). In this model academic achievement is hypothesized to be one causal determinant of academic self-concept. However, it does not preclude a more dynamic model in which subsequent levels of academic achievement and self-concept are each determined by prior levels of achievement and self-concept. According to the path model, math and reading skills are highly correlated with each other while Math and Reading self-concepts are nearly uncorrelated. Reading achievement has a strong, positive direct effect on Reading 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 Reading self-concept. Thus, the I/E model makes many testable predictions besides the lack of correlation between Reading and Math self-concepts. The investigations described below examine empirical support for these predictions.



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

Figure 16. Path Model of Relationships Among Achievements and Reading and Math Self-Concepts: Internal/External Frame of Reference Model

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.

Correlations Between Reading and Math Self-Concepts

Although scores representing Math and Reading self-concepts were derived in each of the original SDQ-I studies (see footnote to Table 8, page 73, for a list of these studies), these correlations are difficult to compare. In a few studies the correlations

were based on unweighted total scores, but in most they were based on factor scores derived from factor analyses that were unique to each study. The earliest SDQ-I research included responses to negatively worded items, though subsequent research demonstrated that these items were biased, and they are no longer included in calculation of the individual scale or total scores of the SDQ-I. To facilitate the comparison of correlations, factor scores derived from the factor analysis described in Chapter 4 (see Table 1, page 31) were used to determine all the Math/Reading correlations for SDQ-I responses from all studies conducted prior to the Marsh (1986d) study (see Table 8).

For the total population the correlation between Math and Reading self-concepts is close to zero (.06), and only 3 of 12 correlations based on individual studies reach statistical significance. The correlations based on the one sample of second-grade students (.49) and the one sample of third-grade students (.46) are substantial, but the correlations vary between -.13 and .17 for the other 10 samples based on responses by fourth-, fifth-, and sixth-grade students. This difference according to grade level is also reflected in the various subtotal correlations. Thus, the correlation across all respondents is .06, but it is .01 for fifth- and sixth-grade students and .17 for second-, third-, and fourth-grade students. Results for males and females considered separately indicate that the lack of correlation is consistent across sex. These findings demonstrate that, with the exception of the youngest children, self-concepts in math and reading are nearly uncorrelated for responses by preadolescents to the SDQ-I.

To test the generalizability of these findings, similar comparisons were made for responses by older subjects. In one large study (Study 8 in Table 8) the SDQ-II was administered to high school students in grades 7 through 12. The Reading/Math correlations did not reach statistical significance at any of the grade levels and was almost zero (-.0002) across all respondents. The SDQ-III has been employed in three studies (Studies 9, 10, and 11 in Table 8) with university students, grade 11 high school students, and a nonstudent population of young adults who were participants in an Outward Bound self-development program. Again, the five Reading/Math correlations were consistently and remarkably close to zero (-.03 to .03), and they did not reach statistical significance for any of the studies.

Correlations between Math and Reading self-concepts have generally been based on responses by students in an academic setting. The importance of the internal comparison process in which self-perceived skills in math and reading are compared and the distinctiveness of the two academic self-concepts may be exaggerated in an academic setting. Hence, the results based on the Outward Bound study are particularly important. This study is based on responses from young adults (ages 16-31, median = 21 years) who were primarily nonstudents and who were participating in a program that emphasized primarily physical outdoor activities and perhaps social relationship skills rather than academic skills. Even for this sample of predominantly nonstudents completing the SDQ-III in a nonacademic setting, support for the relative lack of correlation between Reading and Math self-concepts is strong.

In summary, the results from a variety of studies based on responses from preadolescents, adolescents, and young adults have consistently demonstrated that there is virtually no correlation between Reading and Math self-concepts. The lack of Math/Reading correlation is consistent across age (beyond third grade) and sex and across academic and nonacademic settings. This finding is counterintuitive and contrary to theoretical models, such as the original Shavelson model which postulates that Reading and Math self-concepts combine to form a single, higher order, academic self-concept. However, the findings are consistent with the I/E model and the revision of the Shavelson model and offer support for the validity of interpretation based on the I/E model.

Table 8. Correlations Between Math and Reading/Verbal Self-Concepts in Studies Employing the SDQ-I, SDQ-II, and SDQ-III

	N	Grade	Factor Score Correlations
SDQ-I Studies			
1	305	6	-.02
2(Time 1)	150	4	.08
2(Time 2)	143	4	-.13
3(Time 1)	541	5,6	-.04
3(Time 2)	528	5,6	.06
4	180	4,5,6	.15
5	498	6	.06
6	170	2	.49**
6	103	3	.46**
6	134	4	.01
6	251	5	.17*
7	559	5	-.01
Total Grades 2-4	795		.17**
Total Grades 5-6	2767		.01
Total Males	1970		.10**
Total Females	1592		.06*
Total	3562		.06*
SDQ-II Studies			
8	236	7	-.01
8	223	8	.08
8	181	9	-.05
8	189	10	-.04
8	72	11,12	-.17
Total Males	479		.07
Total Females	422		-.02
Total Grades 7-12	901		.00
SDQ-III Studies			
9	151	University	-.03
10	296	11	-.04
11a	357	Young Adult	-.02
11b	358	Young Adult	-.01
11c	355	Young Adult	.03

Note: For studies 1-7 (the SDQ-I studies in the normative archive), the Math and Reading self-concept scores were derived from a combined factor analysis (see Table 1) so the correlations may differ somewhat from those presented in the original studies cited above. For studies 8 through 11, the Math and Verbal self-concepts were based on factor scores derived from a separate factor analysis of responses from each study.

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

*Studies in the Table include:

- | | |
|---|--|
| (1) Marsh, Parker, and Smith (1983), Study 1. | (7) Marsh, Smith, and Barnes (1985). |
| (2) Marsh, Smith, Barnes, and Butler (1983), Study 1. | (8) Marsh, Parker, and Barnes (1985). |
| (3) Marsh, Smith, Barnes, and Butler (1983), Study 2. | (9) Marsh, Barnes, and Hovevar (1985). |
| (4) Marsh and Groundwater-Smith, unpublished study. | (10) Marsh and O'Niell (1984). |
| (5) Marsh, Relich, and Smith (1983), Study 2. | (11) Marsh, Richards, and Barnes (1986b); a, b, and c represent instruments administered six weeks prior to, at the start of, and at the end of a month-long self-development program called Outward Bound). |
| (6) Marsh, Cairns, Barnes, and Tidman (1984). | |

From Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23, p. 137. Copyright 1986 by the American Educational Research Association. Adapted by permission of the publisher.

The Achievement/Self-Concept Relationship

Though the lack of correlation between Reading and Math self-concepts is consistent with the I/E model, the model may be better examined in studies which include math and verbal achievement scores as well as Math and Reading self-concept measures. Figure 16 illustrates an explicit and counterintuitive pattern of relationships among the four variables representing academic achievements and academic self-concepts in the form of a path model. Results from different studies using the SDQ-I, SDQ-II, and SDQ-III provide a total of 13 analyses to test this path model. Each of these analyses is based on a reanalysis of scores from a previous study. (See footnote in Table 8 for a list of these studies.) The studies employed different math and reading ability measures, including objective test scores, teacher ratings, and school performance. The six analyses which used teacher ratings of achievement occurred at the primary school wherein the same teacher was responsible for teaching both math and reading for each class group; thus, achievement ratings were made by the same person. The achievement tests in Study 7 were administered by the researchers, and those in Study 10 were part of a statewide assessment program.

The high school performance measure in Study 8 was the ability grouping to which each student was assigned on the basis of his or her performance in math and English classes during the previous school year. For year 7, the first year of high school, students were assigned to the same ability group for math and English based on results of a general ability test, so no test of the model was possible. In years 11 and 12, the "ability grouping" was primarily a self-assigned group which reflected student interest and further educational plans, so the use of the ability grouping as an indicator of achievement for this one group may be dubious. Also, historically year 10 has been the typical "school leaving" age which accounts for the smaller sample size even when years 11 and 12 are combined; thus, caution should be used in comparing these results with those of younger students. Since the variables used in these analyses generally are not directly comparable across studies, no attempt was made to estimate the path parameters across different analyses.

Parameter estimates derived for the path model in each of the 13 analyses are presented in Table 9. As predicted by the I/E model, correlations between indicators of verbal and math achievement ($r = .34$) are substantial, ranging from .42 to .94, while correlations between residual measures of Reading/Verbal and Math self-concepts ($r = 12.34$) are much smaller, ranging from -.10 to .19. It is interesting to note that 3 of the 13 estimates of r 12.34 reach statistical significance, that each of these is positive (ranging from .10 to .19), and that each is based on scores from studies using unweighted totals rather than factor scores to represent self-concepts. The other estimates, which are based on factor scores, range from -.10 to .12, and none is statistically significant. This supports earlier contentions that the different areas of self-concept are more clearly differentiated by factor analytically derived scores than by unweighted scores.

The path coefficients representing the relationship between Reading/Verbal self-concept and verbal achievement (p13), and between Math self-concept and math achievement (p24), are positive and statistically significant in all 13 analyses. In dramatic contrast the path coefficients representing the link between math achievement and Reading/Verbal self-concept (p14) and between verbal achievement and Math self-concept (p23) are *negative* and statistically significant in all 13 analyses. For 25 of the 26 parameter estimates. The one exception is a nonsignificant path coefficient for eleventh- and twelfth-grade students in Study 8; however, as mentioned earlier, the use of ability groupings as indicators of achievement may be dubious in this one analysis.

In summary, the parameter estimates in Table 9 provide remarkably strong support for predictions derived from the I/E model. This support is consistent across studies in which the age of the students differs substantially, a variety of indicators of academic achievement are employed, and different self-concept instruments are used.

Marsh (1986d) noted that his support for the I/E model was based entirely on responses by Australian students to one of the SDQ instruments. Therefore, Marsh, Byrne, and Shavelson (in press) tested the I/E model for responses by eleventh- and twelfth-grade Canadian students ($N = 991$) to three different academic self-concept instruments including the SDQ-III. Despite the fact that school performance

Table 9. Path Coefficients for Testing the Internal/External Frame of Reference Model

Study	Achievement Scores	r12.34	p13	p14	p24	p23	r34
1	Teacher Ratings	.00	.53**	-.18*	.52**	-.14*	.42**
2(Time 1)	Teacher Ratings	.17*	.45**	-.13	.39**	-.25*	.60**
2(Time 2)	Teacher Ratings	-.04	.64**	-.29**	.51**	-.23*	.54**
2(Time 2)	Test Scores	-.09	.54**	-.26**	.46**	-.20*	.63**
3(Time 1)	Teacher Ratings	.10*	.47**	-.33**	.44**	-.21**	.76**
3(Time 2)	Teacher Ratings	.19**	.28**	-.10*	.43**	-.17*	.61**
7	Teacher Ratings	.01	.46**	-.33**	.53**	-.12*	.76**
7	Test Scores	.07	.54**	-.19**	.30**	-.21**	.61**
8(Grade 8)	School Performance	.12	.75**	-.58**	.86**	-.71**	.94**
8(Grade 9)	School Performance	-.02	.49**	-.27*	.81**	-.62**	.87**
8(Grade 10)	School Performance	.03	1.03**	-.66**	.73**	-.54**	.88**
8(Grades 11-12)	School Performance	-.10	.31*	.03	.53**	-.45**	.47**
10	Test Scores	-.03	.55**	-.22**	.72**	-.24**	.59**

Note: See note in Table 8 for references to the studies. The label r12.34 is a residual correlation, and the "p's" refer to the standardized path coefficients which are obtained from a multiple regression analysis.

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

From Marsh, H. W. (1986). Verbal and math self-concepts: An internal external frame of reference model. *American Educational Research Journal*, 23, p. 138. Copyright 1986 by the American Educational Research Association. Adapted by permission of the publisher.

measures in mathematics and English were substantially correlated (.51), correlations between Math and Verbal self-concepts were consistently close to zero. The six correlations between Math and Verbal self-concept scales from the various instruments varied from -.05 to .08; the correlation between total scores across the three instruments was .00.

The I/E model predicts that the effect of verbal achievement on Math self-concept, and of mathematics achievement on Verbal self-concept, will be negative. Marsh et al. found support for this prediction for responses from each of the self-concept instruments considered separately and for total scores based on all three instruments.

In subsequent analyses, Marsh et al. expanded the I/E model (see Figure 16) to include total GPA across all school subjects, General-School self-concept, general esteem, and the sex of the subject. However, none of these additional variables altered the general pattern of support for the I/E model.

Inferred Self-Concepts

Results based on the I/E model suggest that both the internal and external comparison processes are operative, and the weights assigned to the two processes

are roughly equal in the formation of Math and Reading self-concepts. An alternative procedure for testing the I/E model is to examine parameter estimates in situations in which one process is expected to be markedly stronger. Applying this approach to data which are available in some of the SDQ-I studies, it was hypothesized that when external observers (e.g., teachers or peers) are asked to infer self-concepts, they rely primarily on externally observable indicators and thus employ primarily the external comparison process.

Research to be described in Chapter 9 indicates that under some circumstances external observers are able to infer self-concepts with a moderate level of accuracy. However, the finding that self-report self-concepts and inferred self-concepts are modestly, or even substantially, correlated does not imply that they are formed in the same way. Though the I/E model was not specifically designed to explain relationships among academic achievement and Math and Reading self-concepts as inferred by others, several observations seem relevant. Previous SDQ-I research on academic self-concepts as inferred by teachers suggests that teacher ratings are primarily a function of their perceptions of a student's actual academic ability. In this sense, their inferred self-concepts reflect the external comparison process rather than the internal comparison process. It is likely that other external observers also emphasize the external comparison process rather than internal comparisons in forming inferred self-concepts. If inferred self-concepts are based only upon an external comparison process, the predicted pattern of parameter estimates for the path model will be quite different. In particular, the correlation between the residual scores for Reading and Math self-concepts is likely to be substantial and positive, and the path coefficients representing the math achievement/Reading self-concept and verbal achievement/Math self-concept links will not be negative.

To examine these predictions, parameter estimates similar to those in Table 9 were determined from those studies in which there were independent estimates of inferred self-concepts and achievement scores in math and reading. Only four studies allowed tests of the predictions (studies in which ratings by the same teacher were used both to infer self-concepts and to estimate academic abilities were not included), and all were based upon preadolescent self-concepts. For two of the cases, self-concepts inferred by teachers were correlated with objective test scores. For the other two cases, self-concepts inferred by peers (another student in the class) were correlated with either teacher ratings of academic ability or achievement test scores. The patterns of parameter estimates for these analyses (see Table 10) differ dramatically from those in Table 9. Correlations between Math and Reading self-concepts as inferred both by teachers and peers are much larger than those based upon self-report measures in the same studies ($r = .47$ to $.58$ vs. $r = -.09$ to $.07$). The path coefficient linking math achievement to Reading self-concept is significantly positive rather than negative for three of the four cases, and the path linking reading achievement to Math self-concept is significantly positive for one case and significantly negative in a second case.

Discussion

The I/E frame of reference model is designed to explain relationships between Reading and Math self-concepts and between these academic self-concepts and corresponding indicators of academic achievement. Development of the I/E model was originally prompted by the observation that Reading and Math self-concepts are relatively uncorrelated with each other, even though verbal and math achievement indicators are substantially correlated with each other and with their corresponding self-concepts. Near-zero correlations between Math and Reading self-concepts were demonstrated in a variety of studies, with correlations of practical significance observed only for second- and third-grade students.

Other I/E model predictions were tested in an examination of the relationships between academic self-concepts and achievement measures. The pattern of relationships between achievement in reading and math and the corresponding measures of self-concept were dramatic and counterintuitive. Despite high correlations between reading and math achievement indicators and the significant correlation of each to the matching measure of academic self-concept, Reading and Math self-concepts were nearly uncorrelated to each other. Furthermore, the direct effect of reading achievement on Math self-concept and the direct effect of math achievement on Reading self-concept were significantly *negative*. This pattern of results was consistent with predictions from the I/E model. According to this model, a high Reading self-concept will be more likely when verbal achievement is high (the external process) and when verbal achievement is higher than math achievement.

Table 10. Path Coefficients for Testing the Internal/External Frame of Reference Model with Inferred Self-Concept Ratings

Study	Inferred By	Achievement Scores	r12,34	p13	p14	p24	p23	r34
2(Time 2)	Teacher	Test Scores	.48**	.25*	.24*	.58**	-.09	.63**
7	Teacher	Test Scores	.58**	.46**	.17**	.36**	.28**	.61**
7	Peers	Teacher Rating	.47**	.21**	.17**	.49**	-.16*	.76**
7	Peers	Test Scores	.49**	.20**	.08	.19**	.04	.61**

Note: See note in Table 8 for references to the studies. The label r12,34 is a residual correlation, and the "p's" refer to the standardized path coefficients which are obtained from a multiple regression analysis.

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

From Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23, p. 138. Copyright 1986 by the American Educational Research Association. Adapted by permission of the publisher.

Thus, once the effect of verbal achievement is controlled for, it is the *difference* between verbal and math achievement that determines Reading self-concept. The direct effect of math achievement is negative in that a higher level of math achievement, given the same level of verbal achievement, will actually lead to a lower level of Reading self-concept. These findings not only demonstrate the clear separation between Math and Reading self-concepts (they are much more distinct than corresponding measures of academic achievement in the two areas), but they also demonstrate that academic self-concepts are affected by different processes than are achievement measures.

In marked contrast to the self-report data, inferred self-concepts based upon peer and teacher responses followed a different pattern of results, and there was no evidence that the internal comparison process was operating. Particularly for teachers, it appears that inferred academic self-concepts reflect little more than their perceptions of objectively defined achievement. Academic self-concepts inferred by teachers are highly correlated with objective achievement measures, but they do not accurately reflect the relativistic nature of self-concepts which is embodied in the external comparison process employed by students in forming their own self-concepts. This suggests that even the external comparison process may not operate the same way in the formation of self-concepts inferred by teachers and those based on students' own self-reports. These findings certainly demonstrate that the formation of one's own self-concepts is affected by different processes than those affecting the self-concepts inferred by significant others.

Research supporting the I/E model and the revision of the Shavelson model (see Chapter 4) demonstrates the extreme separation of Reading and Math self-concepts. This research also makes ambiguous the role and usefulness of general academic

self-concept. General academic self-concept apparently cannot adequately reflect the diversity of specific academic self-concept facets. If the role of academic self-concept is to better understand the complexity of self in an academic context, predict academic behaviors and accomplishments, provide outcome measures for academic interventions, and relate academic self-concept to other constructs, then the specific facets of academic self-concept are more useful than a general academic facet.

Research summarized here emphasizes the separation of Math and Reading self-concepts, but a growing body of research suggests that this separation generalizes to other academic effects as well. Marsh (1984a, 1986c; Marsh, Cairns, Relich, Barnes, & Debus, 1984) demonstrated that the self-attributional pattern that an individual uses to explain academic successes and failures is distinct for verbal and math content areas. Gottfried (1985, 1982) demonstrated the content specificity of intrinsic motivation and anxiety in reading and mathematics. Marsh (1988a), in an analysis of the large, nationally representative High School and Beyond data, showed that attitudes toward math and English classes were nearly uncorrelated even though achievement scores in corresponding content areas were substantially correlated. Daly, Bell, and Korinek (1987) showed that attitudes toward a variety of academic content areas could be explained by higher order factors representing verbal and math content areas that were nearly uncorrelated.

The present application of the I/E model emphasizes academic abilities and self-concepts. However, it is likely that a similar process acts in other areas as well. For example, consider a professional tennis player who is also an excellent golfer and a weekend sports enthusiast who is both an average golfer (which is his or her best sport) and a below-average tennis player. The tennis professional in this example is a better golfer than the weekend sports enthusiast, but may have a self-concept as a golfer which is the same or even poorer; this is consistent with the internal comparison process. Such an internal comparison process may also affect self-concepts in broader areas such as academic versus nonacademic self-concept. Hence, while this application of the I/E model is specific to academic areas, it remains the task of further research to test its application in other areas.

Support of the I/E model and the SDQ-I research upon which it is based has 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 while being accepted by academically poor students. If teachers are able to more accurately infer 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 academically weak students 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 suggests that teachers emphasize absolute measures of academic achievement in inferring academic self-concepts of their students and largely ignore the particular setting which establishes the frame of reference for students' own development of their self-concept. Second, inferred self-concept ratings by teachers (and also peers) overemphasize the external comparison of student academic skills and underemphasize differences in skills in particular academic areas. Thus, a student who is weak in both math and verbal skills but is stronger in one area than the other will tend to have much larger differences in Reading and Math self-concepts than is reflected in the self-concepts inferred by teachers. Finally the I/E model also presents the gratifying prediction that nearly everyone will feel at least reasonably good about himself or herself in at least some areas.

Chapter 8. Self-Attributions for Academic Success and Failure

Self-concept researchers frequently posit a connection between self-concept and self-attributions for one's behavior. Shavelson proposed that self-concept is formed in part by "one's attributions for one's own behavior" (Shavelson & Bolus, 1982, page 3), though this aspect of his definition was not further developed. Other researchers have suggested that the tendency to internalize responsibility is positively correlated with self-concept (e.g., Burns, 1979). This chapter describes the theoretical development of a measure of academic self-attributions, posits a theoretical relationship between dimensions of academic attribution and dimensions of self-concept, and presents empirical tests of theoretical predictions.

Attribution researchers ask subjects for their perceptions of the cause of a particular outcome and examine the perceived causes (i.e., attributions) that people use to explain events. Their arguments for the importance of perceived causes of success and failure have important implications that have been widely applied in educational settings. Individual differences in the way children attribute outcomes to such causes as ability, effort, and luck are related to school performance, self-concept, and academic behaviors (Bar-Tal, 1978; Marsh, Cairns, Relich, Barnes, & Debus, 1984; Weiner, 1979, 1980). In some experimental contexts, attribution researchers have examined individual differences in the way subjects explain their own behavior across different settings (a dispositional or trait emphasis). Other researchers have studied how systematic manipulations in the situation alter attributions (a situational or state emphasis). This chapter focuses on the identification of dispositional tendencies in the way children form attributions about academic outcomes and how these relate to self-concept constructs (see Marsh, 1984a; Marsh, Cairns et al., 1984).

Attribution theorists place emphasis on different causes (e.g., ability, effort, luck) and stress the effects of experimentally manipulated situational variables. For example, Weiner (1972, 1974) argued for a two-dimensional taxonomy in which the bipolar dimensions are locus of control (internal-external) and stability (stable-unstable). Thus, perceived causes can be classified into four types representing 2 x 2 combinations of the endpoints of the two dimensions (i.e., the cause of ability is internal-stable, internal-unstable, external-stable, or external-unstable).

In contrast to the attribution theorists, the study of individual differences in self-attributions stems primarily from "locus of control" research (Rotter, 1966; Stipek & Weisz, 1981) which hypothesizes a generalized expectancy for the internal or external control of events. Some researchers have argued for the need for more specific measures of locus of control. Stipek and Weisz, for example, indicated that achievement measures should be more highly correlated with academic locus of control than with a general measure of locus of control. However, they provided little support for this contention based on existing measures. Marsh, Cairns et al. (1984) also argued that dispositional tendencies should be specific to particular content areas such as academic achievement or even to particular subject areas within an academic setting (e.g., verbal or mathematical achievement). They

attempted to test this assumption with a new instrument, the Sydney Attribution Scale (see below).

Correlations Between SDQ-I and Sydney Attribution Scale

Marsh, Cairns et al. (1984) developed the Sydney Attribution Scale to measure students' perceptions of the causes of their academic successes and failures in reading, math, and general school subjects. The instrument's 18 scales result from the combination of three facets: academic content (math, reading, general school), outcome (success or failure), and perceived cause (ability, effort, or some external cause). Consistent with predictions based on previous self-attribution research, a factor analysis of responses to the Sydney Attribution Scale resulted in seven factors representing attributions: (a) reading ability, (b) math ability, (c) success due to effort, (d) success due to external causes, (e) failure due to lack of ability, (f) failure due to lack of effort, and (g) failure due to external causes. These results demonstrate support for the separation of attributions according to outcome and perceived cause and, at least in the case of ability attributions, for the separation of attributions according to academic content area. Marsh (1984a) then revised the Sydney Attribution Scale by excluding the general school scales and using fewer items to represent the remaining scales, but hypothesized that a similar seven-factor solution would still underlie responses. Conventional/exploratory factor analyses identified these seven factors, and confirmatory factor analyses demonstrated that the seven-factor solution provided a reasonable fit to the data, whereas other plausible models did not.

It is generally posited that the tendency to internalize responsibility is positively related to self-concept, but Marsh, Cairns et al. (1984) argued that this makes sense only for successes. A favorable self-concept is consistent with ability and effort attributions for one's successes but not with a disposition to attribute failure to a lack of effort and particularly not to a lack of ability. Individuals with a high self-concept may be willing to attribute failure to a lack of effort in some specific conditions because a more favorable outcome consistent with a positive self-concept might be expected with more effort. Ability, however, cannot be so easily altered, so it is unlikely that a person with a high self-concept will attribute failure to a lack of ability. Marsh, Cairns et al. also posited that academic attributions, particularly attributions for ability, will be more highly correlated with academic self-concept than with nonacademic self-concept and more highly correlated with self-concepts in the matching academic content area than in nonmatching academic areas.

Based on a research review, Marsh (1984a) proposed and tested the predictions that academic self-concepts will be substantially correlated with success/ability and success/effort attributions (positively) and with failure/ability attributions (negatively), less substantially correlated with failure/effort attributions (negatively), and least correlated with success/external and failure/external scales. He further predicted that the magnitude of correlations will be largest when both self-concept and self-attribution measures are specific to the same academic subject (i.e., reading or math). This pattern of results was expected to generalize to self-concepts inferred by teachers and peers. These predictions emphasize both the convergence of measures which are theoretically or logically connected and the divergence of measures which are designed to assess separate components. Thus, the pattern of results is as important as the actual magnitude of correlations and follows an approach to construct validity which is based upon the logic of multitrait-multimethod analysis.

To replicate and further examine the academic attribution/self-concept relationship, measures from a sample of 559 fifth-grade students were collected to assess dimensions of self-attribution for causes of academic outcomes, multiple dimensions of self-concept, self-concepts inferred by teachers and peers, and academic achievement indicators in math and reading. The results of several aspects of this study (Marsh, 1984a; Marsh, Smith, & Barnes, 1984) have been discussed earlier in this Manual. (Table 3, page 42, shows a factor analysis of the responses to the SDQ-I; Table 7, pages 54-56, and Table 9, page 75, show the correlations between achievement indicators and SDQ-I scales; and Table 10, page 77, shows the relationship between self-concepts and inferred self-concepts.) Material presented here focuses on the relations between multiple dimensions of academic attributions and the SDQ-I scales.

Correlations between responses to the SDQ-I and the Sydney Attribution Scale (Table 11, pages 82-83) support the predicted relationships. The three academic self-concepts and their totals were substantially correlated with success/ability and success/effort scales (positively) and with the failure/ability scale (negatively), less substantially correlated with the failure/effort scale (negatively), and nearly uncorrelated with the two external scales. The four nonacademic self-concepts and their total were less correlated with the self-attribution scores. As predicted, the largest correlations between the self-concept and self-attribution scores occurred between the two reading-specific scores and between the two math-specific scores. These correlations were substantially larger than the other coefficients and clearly support the content specificity of both self-concept and self-attributions.

In addition to the self-concepts based on the student's own self-reports, self-concepts inferred from responses by teachers and peers were also collected. (See Chapter 9 for a discussion of inferred self-concept.) The *sizes* of the correlations between self-attributions and inferred self-concepts were consistently lower than those described above, but the *pattern* of results was very similar (see Table 11). This similarity demonstrates that the predicted pattern of relationships between self-concept and self-attribution is not limited to results based on two self-report instruments completed by the same person. Therefore, alternative explanations for the self-concept/self-attribution relationships based on a response bias or method/halo effect do not seem viable when self-concepts are inferred by external observers.

The results in Table 11 replicate and expand findings by Marsh, Cairns et al. (1984) and demonstrate strong support for a detailed set of predictions based on the earlier research. Seven factors were hypothesized to underlie self-attribution responses to the Sydney Attribution Scale. Conventional/exploratory factor analyses identified these factors, and confirmatory factor analyses demonstrated that the seven-factor solution provided a reasonable fit to the data whereas other plausible models did not. The previous pattern of correlations between self-attributions and self-concepts was replicated, and a similar pattern was identified when self-concepts were inferred from responses by teachers and peers. Support for the predicted pattern of self-attribution/academic achievement relationships was found, and a similar pattern was observed with both academic test scores and teacher ratings of academic ability. In general, students who attribute their academic success to their own ability and, to a lesser extent, to their own effort, tend to have better academic skills and higher academic self-concepts. Students who attribute their academic failures to their lack of ability and, to a lesser extent, to their lack of effort, tend to have poorer academic skills and lower academic self-concepts. Academic self-attributions and academic self-concepts are also specific to particular content areas so that attributions in verbal areas do not generalize to math outcomes. The support for the complicated pattern of predicted relationships between the Sydney Attribution Scale and SDQ-I scores further substantiates the convergent and divergent validity of both instruments, and thus the construct validity of interpretations based upon them.

Table 11. Correlations Between Sydney Attribution Scale, SDQ-I Scales, Inferred Self-Concepts, and Academic Achievement Measures

Criterion	Sidney Attribution Scale Scores ^a													
	Reading		Math		Reading		Math		Success		Failure			
	Ability ¹	Ability ²	Ability ¹	Ability ²	Ability ¹	Ability ²	Ability ¹	Ability ²	Effort	Extr	Ability	Effort	Extr	
Student Self-Concepts														
Physical Appearance	16	20	20	20	21	21	21	21	28	-04	-15	-14	-01	
Physical Abilities	08	15	10	10	18	13	13	13	16	-07	-10	-09	-04	
Peer Relations	17	15	15	15	17	17	17	17	23	-01	-15	-06	-07	
Parent Relations	05	09	09	09	11	06	06	06	19	-04	-09	-06	-10	
Total Nonacademic	17	22	20	20	25	21	21	21	32	-05	-18	-13	-08	
Reading	62	13	57	57	19	37	37	37	23	-19	-38	-28	-04	
Math	09	59	13	13	52	38	38	38	40	-06	-30	-15	-06	
School	24	41	29	29	40	38	38	38	38	-10	-28	-21	-13	
Total Academic	44	52	45	45	51	52	52	52	47	-16	-45	-30	-11	
General	35	42	36	36	41	43	43	41	41	-13	-35	-20	-02	
Self-Concepts Inferred by Peers														
Physical Appearance	-10	-06	-06	-06	-04	-08	-08	-08	04	00	08	05	-07	
Physical Abilities	01	00	02	02	03	-02	-02	-02	03	-05	-03	00	-07	
Peer Relations	-01	00	00	00	02	-04	-04	-04	03	-03	-03	00	02	
Parent Relations	01	01	05	05	03	02	02	02	15	01	-01	-03	-04	
Total Nonacademic	-03	-02	01	01	01	-04	-04	-04	09	-02	01	01	-05	
Reading	28	16	27	27	18	19	19	19	14	-15	-23	-13	-05	
Math	14	32	13	13	27	27	27	27	22	-10	-19	00	-02	
School	13	22	14	14	22	21	21	21	20	-10	-15	-03	-05	
Total Academic	21	27	21	21	26	26	26	26	22	-14	-22	-07	-05	
General	03	00	05	05	02	-02	-02	-02	07	-07	-05	01	-04	
Self-Concepts Inferred by Teachers														
Physical Appearance	08	11	11	11	12	03	03	03	02	-12	-15	-09	-03	
Physical Abilities	08	19	10	10	20	15	15	15	12	-16	-13	-01	-01	
Peer Relations	02	08	07	07	10	01	01	01	01	-13	-03	-05	-09	
Parent Relations	08	06	13	13	05	04	04	04	05	-11	-09	-04	-03	
Total Nonacademic	09	14	15	15	16	08	08	08	08	-18	-15	-07	-08	
Reading	33	24	36	36	25	23	23	23	12	-26	-32	-20	00	
Math	24	38	27	27	38	26	26	26	18	-26	-34	-19	-06	
School	20	32	31	31	33	25	25	25	16	-26	-34	-18	-04	
Total Academic	30	34	34	34	35	27	27	27	17	-28	-36	-21	-04	
General	12	24	24	24	26	16	16	16	09	-21	-28	-17	-03	

Note: Decimal points are omitted. Correlations with an absolute value greater than .12 are statistically significant ($p < .01$, two-tailed).

^aTwo variations of the reading and math scales for the Sydney Attribution Scale were considered, where Reading Ability¹ = RSA - RFA, Math Ability¹ = MSA - MFA, Reading Ability² = RSA + RSE - RSX - RFA - RFE, and Math Ability² = MSA + MSE - MSX - MFA - MFE (see Marsh, 1984a, for greater detail).

Table 11. (continued) Correlations Between Sydney Attribution Scale, SDQ-I Scales, Inferred Self-Concepts, and Academic Achievement Measures

Criterion	Sidney Attribution Scale Scores ^a													
	Reading		Math		Reading		Math		Success		Failure			
	Ability ¹	Ability ²	Ability ¹	Ability ²	Ability ¹	Ability ²	Ability ¹	Ability ²	Effort	Extr	Ability	Effort	Extr	
Achievement Test Scores														
Reading	41	25	37	37	25	27	27	27	01	-25	-38	-21	06	
Math	20	33	20	20	33	22	22	22	09	-24	-30	-12	06	
Teacher Ability Ratings														
Reading	35	24	36	36	26	26	26	26	11	-29	-33	-17	02	
Math	22	38	25	25	38	28	28	28	16	-28	-32	-14	-01	
School	29	34	32	32	34	28	28	28	15	-29	-35	-16	-03	
Multiple R Based on Five Achievement Indicators														
Multi R	44	42	41	41	41	32	32	32	19	32	41	22	12	

Note: Decimal points are omitted. Correlations with an absolute value greater than .12 are statistically significant ($p < .01$, two-tailed).

^aTwo variations of the reading and math scales for the Sydney Attribution Scale were considered, where Reading Ability¹ = RSA - RFA, Math Ability¹ = MSA - MFA, Reading Ability² = RSA + RSE - RSX - RFA - RFE, and Math Ability² = MSA + MSE - MSX - MFA - MFE (see Marsh, 1984a, for greater detail).

The Self-Serving Effect

Individuals are more likely to attribute their own success to internal causes such as ability and effort and to attribute their failure to external causes such as task difficulty, luck, and the influence of powerful others. In a review of the research, Zuckerman (1979) reported that of 38 studies, 27 (71.0%) found subjects taking more responsibility for success than for failure, while two (5.3%) found subjects accepting more responsibility for failure than for success. This finding, sometimes called "self-serving bias" or "hedonic bias" is labeled the "self-serving effect" in this chapter.

Most attribution researchers (e.g., Bradley, 1978; Harvey & Weary, 1984; Snyder, Stephan, & Rosenfield, 1978; Zuckerman, 1979) hypothesize that the motivation to take credit for success and to deny responsibility for failure protects or enhances self-esteem. Riess, Rosenfield, Melburg, and Tedeschi (1981) suggest that the self-serving effect could represent either conscious, intentional distortions that protect one's self-esteem, or unconscious, unwitting distortions in perceptions of causality that accurately reflect one's self-perceptions. Alternatively, Miller and Ross (1975) proposed an information processing hypothesis in which the self-serving effect is explained by nonmotivational influences. Whereas a conscious distortion of self-perceptions clearly represents a bias in self-attributions, the other explanations may not. Instead, as suggested by Riess et al., "they imply that individuals actually perceive themselves as more responsible for their positive than for their negative outcomes and accurately report their true private perceptions when offering causal attributions for these outcomes" (page 225).

Though many studies of the self-serving effect consider only attributions summarized by a single internal/external score, some have examined attributions to specific causes (e.g., ability, effort, task difficulty, and luck). Such studies have generally found substantial self-serving effects for ability and effort attributions but much smaller or nonsignificant self-serving effects for attributions to external causes (Marsh, 1986c). Thus, the size of the self-serving effect varies systematically with the perceived cause. However, this pattern is obscured when researchers combine responses to different perceived causes to form an overall internal/external score or employ rating tasks that force an artificial interdependency among the perceived causes.

In summary, there is wide support for the existence of a self-serving effect, though it probably occurs primarily with ability and effort attributions. However, there is considerable controversy about how the self-serving effect should be interpreted. Different researchers have identified both motivational and nonmotivational components of the self-serving effect. Ultimately, the explanation of the self-serving effect as entirely either a motivational bias or a valid representation of self-perceptions must be overly simplistic. Consistent with Heider's (1958) original formulation, attributions are probably a function of both objective information and motivational tendencies.

In most previous research there was little interest in the size of the self-serving effects for different respondents. Individual differences were considered a source of error. However, it is likely that there are systematic individual differences in the size of self-serving effects, and these dispositional tendencies will be related to other individual difference characteristics. The remainder of this chapter presents an investigation into individual differences. A further analysis is presented of two studies discussed earlier in this chapter (Marsh, 1984a; Marsh, Cairns et al., 1984) and a study by Thomas (1984) in which separate measures for reading and mathematics were available for self-attributions for academic successes and failures, multiple dimensions of self-concept, and academic achievement indicators.

The purposes of the studies were to examine how the size of the self-serving effect varies with the particular cause being considered, the extent to which it generalizes across different academic content areas, and how it is related to the respondent's level of academic self-concept and achievement.

In Study 1, 226 fifth-grade students (primarily 10-year-olds) completed the Sydney Attribution Scale, the SDQ-I, and two standardized measures of reading achievement (see Marsh, Cairns et al., 1984). In Study 2, 559 fifth-grade students completed the Sydney Attribution Scale, the SDQ-I, and standardized measures of reading and mathematics. In addition, classroom teachers judged the children's reading and mathematics abilities and inferred self-concepts in the areas covered by the SDQ-I scales for each student. In Study 2, inferred self-concept ratings were also collected from students' classmates (see Marsh, 1984a; Marsh, Smith, & Barnes, 1985). In Study 3, 122 ninth-grade students completed the Sydney Attribution Scale and the SDQ-II. Although no standardized measures of academic achievement were collected, all students were assigned to classes in mathematics and/or English according to their previous academic performance in these classes, and in the absence of better measures of achievement, this ability grouping was used (see Marsh, 1986c).

Effect of Outcome on Academic Attributions

Demonstration of a self-serving effect requires that attributions for the perceived causes of success outcomes be more internal than those of failure outcomes. In the ANOVAs in Table 12, page 86, the self-serving effect is represented by the outcome effect — the difference between success and failure attributions. Interactions involving the outcome effect demonstrate ways in which the self-serving effect depends on other variables. For all three studies the self-serving effect is strong; the outcome effect is large, and attributions are more internal for success than for failure (see results for the total groups in Figure 17, page 87). However, the self-serving effect varies significantly and substantially with both the perceived cause and ability level. The outcome x cause interaction is shown in Figure 17. The self-serving effect was largest for ability attributions, slightly smaller for effort attributions, and much smaller or nonexistent for external attributions. In all three studies the self-serving effect was significantly larger for ability attributions than for effort attributions, and the self-serving effect for external scales was small (not even statistically significant in Study 1). Thus, whereas the self-serving effect was strong for attributions of ability and effort, it was weak or nonexistent for attributions to external causes.

Figure 17 also illustrates the outcome x achievement interaction (see results for high- and low-ability groups). The self-serving effect was substantially larger for students with the highest levels of achievement and smaller for students with the lowest levels of achievement. The results for the two middle levels of achievement were also consistent with this trend, though separate graphs for these groups are not presented. As shown in Figure 17, subsequent analyses of all three studies demonstrated that high-ability students were significantly more internal in their attributions for success outcomes than were low-ability students and that they were significantly more external in their attributions about failure outcomes than were low-ability students.

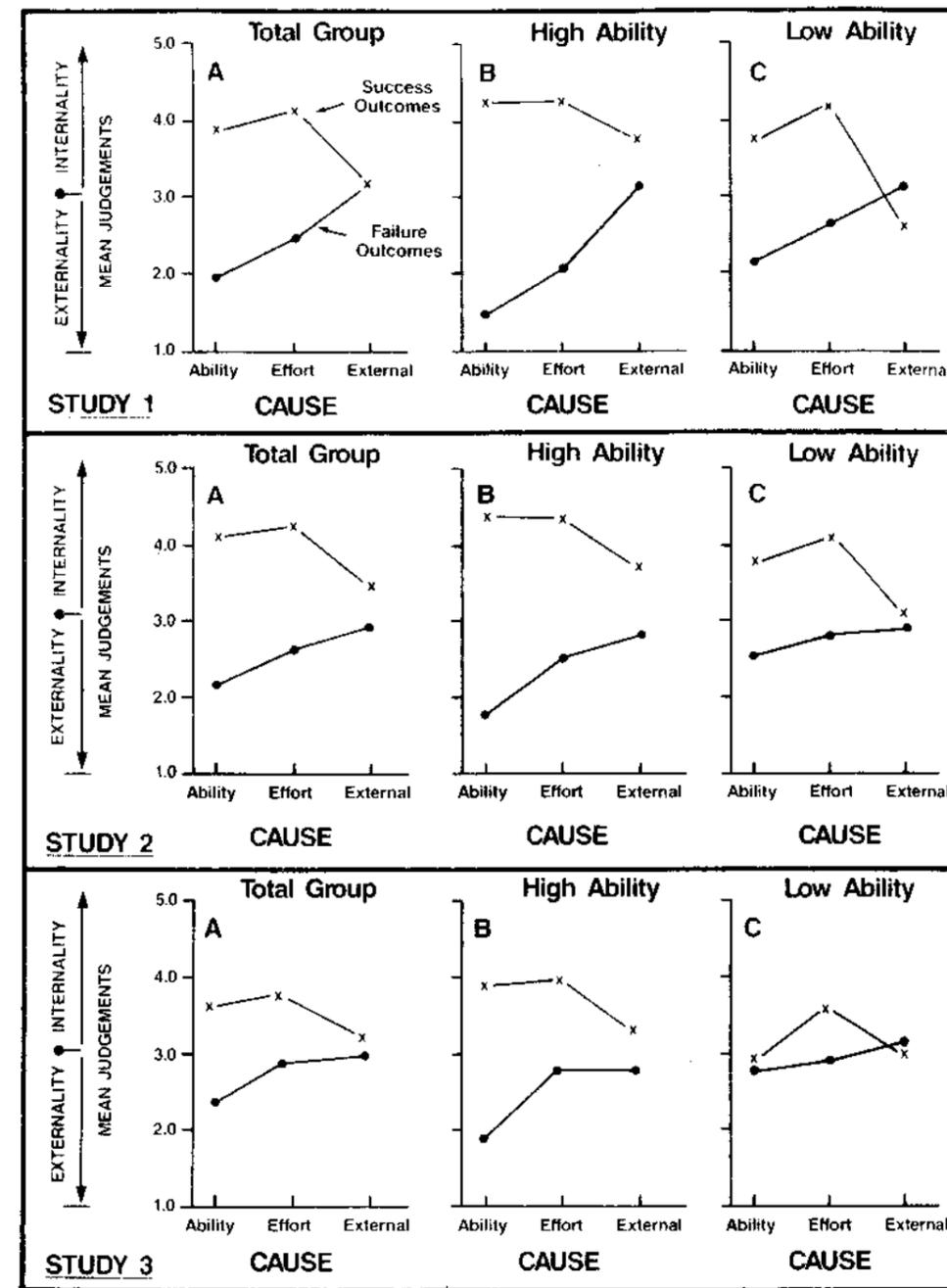
This discussion of the self-serving effect has emphasized the similarity in the three studies. However, the size of the self-serving effect appears to be smaller for responses by the high school students (Study 3) than for primary school students (Studies 1 and 2). This difference is apparent in Figure 17 and in the fact that outcome accounted for a much larger percentage of the variance in Studies 1 and 2 (29.6% and 31.1%, respectively; see Table 12) than in Study 3 (16.5%). In fact, there was no self-serving effect for ability attributions in the responses of the low-ability students in Study 3. The smaller effect in Study 3 suggests that the self-serving

Table 12. Effect of Level of Achievement and the Three Sydney Attribution Scale Facets on the Internality/Externality of Academic Self-Attributions

Effect	Study 1 (N = 226)			Study 2 (N = 559)			Study 3 (N = 122)		
	SS	df	F	SS	df	F	SS	df	F
Total	3297.8			9309.0			1326.0		
Between subjects									
Level of Achievement (L)	1.4	3	0.82	3.7	3	0.97	5.1	3	1.43
Within subjects									
Outcome (O)	977.5	1	791.60**	2,896.9	1	1327.42**	219.1	1	140.04**
Content Area (C)	1.1	1	5.11	42.9	1	121.56**	13.2	1	48.69**
Perceived Cause (P)	72.2	2	40.89**	124.7	2	52.86**	34.4	2	20.62**
O x L	78.4	3	21.16**	175.4	3	26.79**	52.3	3	11.14**
C x L	1.9	3	2.81	2.1	3	1.97	0.7	3	.91
P x L	43.3	6	8.18**	36.8	6	5.20**	3.3	6	.65**
O x C	.6	1	0.11	41.9	1	56.49**	16.6	1	19.36**
O x C x L	4.7	3	3.13	5.6	3	2.97	1.1	3	.44
O x P	502.1	2	239.99**	615.3	2	356.77**	64.8	2	63.69**
O x P x L	10.4	6	1.67	33.4	6	6.45**	24.4	6	7.98**
C x P	20.0	2	56.89**	6.06	2	12.85**	3.9	2	10.57**
C x P x L	1.3	6	1.27	1.5	6	1.05	0.9	6	.84
O x C x P	37.6	2	51.76**	8.3	2	10.85**	3.3	2	5.71*
O x C x P x L	1.0	6	0.46	5.6	6	2.42	1.8	6	1.02

Note: For each effect the corresponding mean square error term = (SS/df) (F ratio). *p < .01; **p < .001.

From Marsh, H. W. (1986). The self-serving effect (bias?) in academic attributions: Its relations to academic achievement and self-concept. *Journal of Educational Psychology*, 78, p. 194. Copyright 1986 by the American Psychological Association. Reprinted by permission of the publisher.



Scores vary along the actual 1- to 5-point scale used for the attribution instrument. Responses to the external scale were reversed so that higher scores reflect more internal-responses. Note: A = Total group; B = Students with the highest ability; C = Students with the lowest ability.

Figure 17. Outcome x Cause Interaction: High and Low Ability Students from the Total Group

From Marsh, H.W. (1986). The self-serving effect (bias?) in academic attributions: Its relations to academic achievement and self-concept. *Journal of Educational Psychology*, 78, p. 194. Copyright 1986 by the American Psychological Association. Reprinted by permission of the publisher.

effect and its relation to academic achievement may be age related, though the many differences among the three studies dictate that such an interpretation must be made cautiously.

Individual Differences

The results of most research on the self-serving effect and the present findings demonstrate that, when averaged across all respondents, attributions for success are more internal than are attributions for failures. The purpose of a second set of analyses was to determine how individual differences in the size of the self-serving effect vary with other individual difference characteristics. A set of six difference scores (differences between attributions for success and failure outcomes) was used to infer the size of the self-serving effect for each individual regarding the perceived causes in each content area. These six self-serving effect indicators were then correlated with each other and with measures of self-concept, self-concepts inferred by teachers and peers, and academic achievement (see Table 13).

Correlations with Self-concept Scores. Correlations between the six self-serving effect indicators and the multiple self-concepts form a systematic pattern of relations that is reasonably consistent across all three studies, although such comparisons are complicated by the different self-concept factors considered in each study. Across all three studies, the self-serving effects for the ability and effort scales were not substantially correlated with nonacademic self-concepts (mean $r = .20$). In each study the highest correlations were between the self-serving effects for ability scales in a particular content area and the academic self-concept in the same content area (mean $r = .64$). Thus, the self-serving effects for math ability scales were highly correlated with Math self-concept ($r = .71, .61, \text{ and } .71$) but not with Reading self-concept ($r = .22, .18, \text{ and } .27$). The self-serving effects for the Reading ability scales were highly correlated with Reading self-concept ($r = .57, .63, \text{ and } .64$) but not with Math self-concept ($r = .09, .14, \text{ and } .24$). In contrast, external attributions — particularly in Studies 1 and 2 — were less highly correlated with the self-concept responses. In summary, in each of the three studies, self-serving effects representing the ability and effort scales were substantially correlated with academic self-concepts. In particular, the self-serving effects based on attributions of ability were quite content specific.

Correlations with Inferred Self-concepts. Inferred self-concepts — those inferred by the classroom teachers and by peers — were collected only in Study 2. The pattern, if not the magnitude, of correlations with the self-serving effect indicators was similar to that observed with the self-concept scores. Correlations were modest between self-serving effects based on external attributions and all inferred self-concepts (mean $r = .08$), as were correlations between the self-serving effects based on ability and effort attributions and inferred self-concepts in nonacademic areas (mean $r = .05$). In contrast, self-serving effects based on ability and effort attributions were more substantially correlated with inferred self-concepts in academic areas (mean $r = .21$). In particular, for self-concepts inferred by both teachers and peers, the highest correlations were between the reading ability attributions and inferred Reading self-concepts (mean $r = .30$) and between the math ability attributions and the inferred Math self-concepts (mean $r = .35$). Hence, the correlations between the self-serving effect indicators and the inferred self-concepts were also content specific.

Correlations with Academic Achievement Indicators. In Study 1 the self-serving effect indicators and the reading achievement scores were positively correlated. The self-serving effect based on reading ability attributions was more highly correlated with reading achievement (.35) than was the self-serving effect based on math ability attributions (.21). In Study 3 the correlation between the math

Table 13. Correlations Between the Self-Serving Effect and Self-Concepts, Inferred Self-Concepts, and Achievement Scores

Measure	Study	Self-Serving Effect Indicators: Differences in Perceived Causes					
		MSA-MFA	MSE-MFE	MSX-MFX	RSA-RFA	RSE-RFE	RSX-RFX
Self-Serving Effect Indicators MSE-MFE	1	.63**					
	2	.63**					
	3	.64**					
MSX-MFX	1	-.10	-.20**				
	2	.14**	.05				
	3	.29**	.18*				
RSA-RFA	1	.35**	.46**	-.11			
	2	.38**	.44**	.07			
	3	.36**	.14	.11			
RSE-RFE	1	.40**	.72**	-.19**	.67**		
	2	.33**	.59**	.04	.64**		
	3	.35**	.40**	.23*	.57**		
RSX-RFX	1	-.12*	-.20**	.55**	-.09	-.25**	
	2	.20*	.15**	.53**	.29**	.19**	
	3	.20*	.04	.39**	.51**	.33**	
Self-Concepts ^a Math	1	.71**	.49**	-.07	.09	.26**	-.11
	2	.61**	.46**	-.02	.14**	.22**	.03
	3	.71**	.41**	.06	.24*	.24*	.12
Reading	1	.22**	.36**	-.03	.57**	.54**	-.10
	2	.18**	.25**	.05	.63**	.43**	.18**
	3	.27**	.11	-.03	.64**	.43**	.22*
School	1	.50**	.49**	-.14*	.25**	.39**	-.11
	2	.46**	.41**	-.04	.28**	.29**	.08
	3	.58**	.31**	.13	.56*	.41**	.28**
Appear	1	.33**	.15*	-.06	.14*	.05	-.08
	2	.26**	.30**	-.01	.20**	.25**	.07
	3	.20*	.01	.15	-.02	-.08	-.04
Physical	1	.25**	.16*	-.16*	.27**	.14*	-.18**
	2	.21**	.21**	.02	.12*	.15**	.03
	3	.17	.08	.11	.23*	.11	.23*

Note: Each attribution score is indicated by three letters which represent content (M = Math, R = Reading), outcome (S = Success, F = Failure), and attribution (A = Ability, E = Effort, X = External), respectively. The responses to the external scales have been valenced so that larger values represent more internal responses. The attribution variables used in this analysis are the differences between responses to the success and failure outcomes. For example, the first attribution variable is the difference between ability attributions in mathematics for successes and for failures.

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

^aSince different self-concept instruments were used in each study, some of the scales appeared in only one or two of the three studies.

^bSelf-concepts inferred by teachers (T) and by peers (P) were collected only as part of Study 2.

^cThe Total score in Study 2 is the sum of the achievement test scores and the teacher ratings.

^dStream scores were collected only in Study 3, and English stream scores were available only for the students from the boys' school in that study.

^eSex was coded so that negative correlations indicate that males scored higher than females.

Table 13. (continued) Correlations Between the Self-Serving Effect and Self-Concepts, Inferred Self-Concepts, and Achievement Scores

Measure	Study	Self-Serving Effect Indicators: Differences in Perceived Causes					
		MSA-MFA	MSE-MFE	MSX-MFX	RSA-RFA	RSE-RFE	RSX-RFX
Self-Concepts ^a	1	.26**	.24**	-.21**	.25**	.22**	-.11
	2	.21**	.22**	-.04	.21**	.18**	.00
Parents	1	.19**	.21**	-.17**	.28**	.31**	-.12
	2	.15**	.18**	-.05	.10	.16**	-.02
	3	.23*	.25**	-.01	.17	.23*	.19*
Same Sex	3	.31**	.17	-.07	.30**	.33**	.15
Opposite Sex	3	.27**	.10	.05	.03	-.04	-.09
Emotional	3	.32**	.21*	.05	.21*	.16	.20*
Honesty	3	.34**	.32**	.10	.25**	.41**	.11
General	2	.37**	.38**	-.01	.31**	.31**	.12**
	3	.41**	.23**	.04	.27**	.19*	.25**
Inferred Self-Concepts ^b	2(T)	.37**	.24**	.14**	.24**	.18**	.20**
	2(P)	.31**	.17**	.05	.14**	.06	.08
Reading	2(T)	.24**	.15**	.19**	.33**	.24**	.25**
	2(P)	.16**	.14**	.05	.27**	.18**	.12**
School	2(T)	.32**	.20**	.16**	.27**	.21**	.22**
	2(P)	.22**	.16**	.06	.13**	.09*	.05
Appear	2(T)	.11*	.06	.07	.08	.08	.10*
	2(P)	-.06	-.01	-.03	-.10*	-.03	-.07
Physical	2(T)	.19**	.10*	.12*	.08	.04	.14**
	2(P)	.00	.03	-.02	.01	.00	.00
Peers	2(T)	.08	.04	.03	.02	.03	.06
	2(P)	.00	.02	.04	-.01	.00	.04
Parents	2(T)	.06	-.01	.07	-.08	.11*	.08
	2(P)	.01	.09*	-.01	.01	.11*	-.06
General	2(T)	.24**	.15**	.14**	.20**	.17**	.17**
	2(P)	.00	.01	.04	.03	.04	.02

Note: Each attribution score is indicated by three letters which represent content (M = Math, R = Reading), outcome (S = Success, F = Failure), and attribution (A = Ability, E = Effort, X = External), respectively. The responses to the external scales have been valenced so that larger values represent more internal responses. The attribution variables used in this analysis are the differences between responses to the success and failure outcomes. For example, the first attribution variable is the difference between ability attributions in mathematics for successes and for failures.

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

^aSince different self-concept instruments were used in each study, some of the scales appeared in only one or two of the three studies.

^bSelf-concepts inferred by teachers (T) and by peers (P) were collected only as part of Study 2.

^cThe Total score in Study 2 is the sum of the achievement test scores and the teacher ratings.

^dStream scores were collected only in Study 3, and English stream scores were available only for the students from the boys' school in that study.

^eSex was coded so that negative correlations indicate that males scored higher than females.

Table 13. (continued) Correlations Between the Self-Serving Effect and Self-Concepts, Inferred Self-Concepts, and Achievement Scores

Measure	Study	Self-Serving Effect Indicators: Differences in Perceived Causes					
		MSA-MFA	MSE-MFE	MSX-MFX	RSA-RFA	RSE-RFE	RSX-RFX
Achievement Indicators							
Test Scores							
Reading	1	.21**	.14*	.35**	.35**	.17**	.39**
	2	.25**	.13**	.21**	.41**	.17**	.28**
Math	2	.33**	.18**	.22**	.20**	.07	.26**
Teacher Ratings							
Reading	2	.24**	.14**	.21**	.35**	.20**	.29**
	2	.38**	.21**	.20**	.22**	.13**	.25**
Math							
Total ^c							
Reading	2	.27**	.16**	.21**	.46**	.23**	.32**
	2	.41**	.25**	.22**	.26**	.15**	.28**
Math							
Stream Scores ^d							
Math	3	.48**	.21*	.17	.45**	.15	.23*
	3	.37**	.20*	.03	.44**	.05	.24*
English							
Sex ^e							

Note: Each attribution score is indicated by three letters which represent content (M = Math, R = Reading), outcome (S = Success, F = Failure), and attribution (A = Ability, E = Effort, X = External), respectively. The responses to the external scales have been valenced so that larger values represent more internal responses. The attribution variables used in this analysis are the differences between responses to the success and failure outcomes. For example, the first attribution variable is the difference between ability attributions in mathematics for successes and for failures.

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

^aSince different self-concept instruments were used in each study, some of the scales appeared in only one or two of the three studies.

^bSelf-concepts inferred by teachers (T) and by peers (P) were collected only as part of Study 2.

^cThe Total score in Study 2 is the sum of the achievement test scores and the teacher ratings.

^dStream scores were collected only in Study 3, and English stream scores were available only for the students from the boys' school in that study.

^eSex was coded so that negative correlations indicate that males scored higher than females.

stream scores and the self-serving effects based on math ability attributions was substantial (.48) but only slightly higher than the correlation with self-serving effects based on reading ability attributions (.45). Similarly, reading stream scores were substantially correlated with the self-serving effect based on reading ability attributions (.44) and almost as highly correlated with the self-serving effect based on math ability attributions (.37).

Study 2 provides a much stronger basis for examining the relations among the set of self-serving effects and academic achievement in different content areas because it contains achievement indicators for reading and mathematics based on test scores and teacher ratings. Here, reading achievement was most highly correlated with the self-serving effect based on reading ability attributions (.46 for the total score), while mathematics achievement was most highly correlated with the self-serving effect based on mathematics ability attributions (.41 for the total score). This same pattern, both for reading and math scores, occurred with objective test scores, with teacher ratings, and their total. Thus, the self-serving effects were moderately correlated with academic achievement indicators, and at least the self-serving effects based on attributions of ability were content specific.

Summary of Individual Differences. In summary, the six self-serving effect indicators form a systematic and logical pattern of relationships with self-concept, inferred self-concepts, and academic achievement. They are significantly correlated with *academic* self-concepts, with inferred *academic* self-concepts, and with academic achievement. Furthermore, self-serving effects, particularly those based on ability attributions, are content specific. This content specificity is most evident in the correlations with self-concepts, but it is also evident in the correlations with different academic achievement indicators and with self-concepts inferred by teachers and peers.

Discussion

Consistent with previous research, the results of this study demonstrate that self-attributions for the perceived causes of success are more internal than those for failure. However, the findings also demonstrate that the size of the self-serving effect depends upon the particular cause that is being evaluated, the individual characteristics of the person making the attributions, and the content area in which the attributions are being made.

The dependency of the self-serving effect on the particular cause has not typically been reported by other researchers because they usually collapse responses from different causes to form a single internal-external score. However, this dependency is consistent with findings from eight studies described by Marsh (1986a). In those studies the self-serving effect was also large for ability and effort scales but was smaller or nonexistent for external scales (task difficulty and luck).

The dependency of the self-serving effect on ability level has not been previously emphasized. In the present study this dependency was demonstrated in an initial ANOVA and in subsequent examination of correlations. Both analyses showed that students who are more academically able are more likely to attribute their academic successes internally and their academic failures externally than are students who are less academically able. Furthermore, in Study 2, and to a lesser extent in Study 3 where multiple indicators of achievement in reading and mathematics were available, the effects of achievement on self-serving effects inferred from ability scales were content specific. Students who were more able in reading had larger self-serving effects for attributions of ability in reading, and students who were more able in mathematics had larger self-serving effects for attributions of ability in mathematics. This systematic and logical pattern of relations between the self-

serving effects and the achievement indicators cannot reasonably be explained as a motivational response bias (e.g., Bradley, 1978) but should be interpreted as an informational influence on the self-serving effect (e.g., Miller & Ross, 1975). The pattern of relations also provides support for the construct validity of the academic attributions.

The dependency of the self-serving effect on self-concept is also systematic and logical, but alternative explanations for this relationship exist. According to a "validity interpretation," it is reasonable that students with high academic self-concepts should attribute academic success internally and academic failure externally. To attribute success externally or to attribute failure internally would be inconsistent with their high academic self-concept. Thus, the positive correlation between academic self-concepts and the self-serving effect in academic attributions is predictable and offers support for the construct validity of both self-concept and self-attribution. However, according to a "bias interpretation" in which the self-serving effect is viewed as a motivational bias in the way subjects respond to the academic attribution instrument, it is reasonable that a similar motivational bias would affect the self-concept responses; this would cause responses to the two measures to be positively correlated.

A more detailed examination of the two alternatives provides further support for the validity interpretation. First, the validity interpretation is clearly consistent with the finding that self-serving effects based on academic attributions are more highly correlated with academic self-concepts than with nonacademic self-concepts, but the bias interpretation probably is not. Second, the extreme content specificity of the correlations for self-serving effects based on ability attributions and matching areas of academic self-concept is consistent with the validity interpretation, but apparently not with the bias interpretation. Third, the bias interpretation is inconsistent with the positive correlations between the self-serving effect indicators and self-concepts inferred by teachers and peers. In particular, the content specificity of the relations between inferred self-concepts and self-serving effect indicators — two sets of measures completed by different individuals — was similar to that observed with the self-concept scores — two sets of measures completed by the same individual. Finally, and perhaps most importantly, the validity interpretation is consistent with the dependency of the self-serving effects on academic achievement, while the bias interpretation is not. Thus, there is stronger support for the validity interpretation of the self-concept/self-serving effect relationship than for the bias interpretation.

In summary, individual differences in the size of the self-serving effect are logically related to individual differences in academic self-concepts, to academic self-concepts inferred by significant others, and to academic achievement. For example, students who are particularly able at mathematics have high Math self-concepts and are inferred by teachers and peers to have high Math self-concepts. They are more likely to attribute success in mathematics to their ability and less likely to attribute failure in mathematics to their lack of ability than are students with poor mathematical abilities, poor Math self-concepts, and poor inferred Math self-concepts. The pattern of relationships is most clear for attributions of ability, but it is reasonable that ability attributions should be most strongly related to academic achievement and self-concept. Although there is ample evidence from other research to demonstrate that motivational biases can influence the self-serving effect, it seems unreasonable to interpret the effect of self-concept, and particularly the effect of achievement, as motivational response biases in this study. The argument that the effect of self-concept is not a motivational bias in this study is particularly important since most interpretations of the self-serving effect as a response bias assume that the purpose of the bias is to protect or enhance self-concept. Ego-provoking manipulations in other studies apparently result in response biases for self-attributions in some situations. However, in this study positive correlations

between the self-serving effect and self-concept apparently represent a logical and reasonable way to infer causality that is not motivated by the need to distort attributions.

The general consistency of the results across the three studies provides support for the generalizability of the findings. Nevertheless, there appear to be some differences in results based on preadolescent and adolescent responses. For the older students in Study 3, the self-serving effects for reading were less correlated with self-serving effects for mathematics than for the younger students, so that the self-serving effects may become more content specific with age. The size of the self-serving effect also appears to vary with age. For the preadolescents, even those with the lowest ability levels, there was a substantial self-serving effect for ability attributions. In contrast, the self-serving effect was smaller for the older children; for those with the lowest achievement levels, the self-serving effect based on ability attributions completely disappeared. This suggests that as less-able children grow older and continue to receive consistently negative feedback about their achievement levels, high ability becomes less viable as an explanation of academic success, and poor ability becomes more viable as an explanation of academic failure. It is also interesting to note that even for the least able, older subjects, the self-serving effect for effort attributions was still substantial. This is consistent with the interpretation of effort as a controllable cause, whether or not the effort is typical or atypical. Although the differences in design and instruments make comparisons across the studies tenuous, the findings suggest that there may be a developmental trend in the self-serving effect and particularly in the self-serving effect based on ability attributions.

Summary and Implications

Results summarized in this chapter demonstrate a predictable and consistent pattern of relations between multidimensional self-concepts and multidimensional attributions for the causes of academic success and failure, and between these self-report measures and academic achievement indicators. Thus, these results provide support for the construct validity of both multidimensional self-concepts based on responses to the SDQ-I and multidimensional self-attributions based on the Sydney Attribution Scale. The results also support the contention by Shavelson et al. (1976) that self-attributions for one's own behavior are part of the basis for forming self-concept.

The predicted pattern of self-concept/self-attribution relationships was based on the assumption that students form their academic self-attributions in a way that is consistent with their academic self-concepts. However, a similar pattern of predictions is postulated by self-worth theory (Covington, 1984; Covington & Omelich, 1979), which hypothesizes that self-worth is determined by self-attributions. The basic premise of the theory is that students formulate self-attributions and behavior so as to lead to feelings of self-worth, a notion that is consistent with the motivational basis of the self-serving effect. In fact, there may not be any contradiction in the two approaches, which differ primarily in the causal ordering of the self-concept and self-attribution constructs; it is likely that changes in self-concept will lead to changes in self-attribution and that changes in self-attributions will lead to changes in self-concept. Marsh (1984a), in speculating on this possibility, proposed that academic self-concept, academic self-attributions, and academic achievement are interwoven in a network of reciprocal relationships which form a dynamic equilibrium so that a change in any one will produce changes in the others in order to reestablish a new equilibrium. Also consistent with this perspective is the suggestion that changes in academic self-concept are caused by and cause changes in academic achievement.

Chapter 9. Inferred Self-Concepts

Self-concept ratings by others, called "inferred self-concepts," are used to determine how accurately self-concept can be inferred by external observers, to validate interpretations of responses to self-concept instruments, and to test a variety of theoretical hypotheses. Shavelson specifically postulated that self-concept is influenced by the evaluations of significant others, but he also emphasized that self-concept measured by self-report is a separate construct from self-concept inferred by external observers. In a series of SDQ studies discussed in this Chapter, significant others were asked to infer the self-concepts of students who had completed an SDQ instrument, and multitrait-multimethod analyses were used to examine self-other agreement.

Theoretical Basis

Symbolic interactionists argue that self-concept emerges from a person's social interaction with others, that self-concept is based on the ways others respond to him or her, and that a person's perceptions of others' responses reflect their actual responses (Kinch, 1963). Although the relationship between self-perceptions and the perceived perceptions of others is a critical issue for symbolic interactionists, Shrauger and Schoeneman (1979; also see Kinch, 1963; Miyamoto & Dornbusch, 1956) argue that support for the theory also requires a "congruence between (a) self-perceptions and others' actual perceptions of the person and (b) perceived other-evaluations and actual other-evaluations" (p. 552). Thus, although self-other agreement is consistent with the symbolic interactionist perspective, the predicted strength of the relationship depends on the accuracy of perceived other-evaluations and on the strength of the relationship between self-perceptions and the person's evaluation of perceptions by others.

Research involving ratings by others is plagued with a variety of ways in which the rating task can be formulated. For example, external observers can be asked what *they* think or feel about a person. On the other hand, external observers can be asked to use their observations to infer what that person thinks about himself or herself (i.e., inferred self-concept). The first approach might be appropriate to determine how accurately a person views himself or herself compared to the perceptions of others. However, because self-concept is based upon self-perceptions, whether accurate or not, the second approach is used in most self-concept research (see Wells & Marwell, 1976, pp. 136-142, for further discussion). Nevertheless, this distinction is not always clear in self-concept research, and perhaps not to the external observers even when researchers ask for inferred self-concepts. In the SDQ research described in this chapter, ratings by others refer to inferred self-concepts.

Even among researchers who agree that ratings by others should be inferred self-concept ratings, there is disagreement about their relevance. At one extreme Combs, Soper, and Courson (1963) argue that self-report measures of self-concept are unduly affected by sources of bias and that inferred ratings by external observers provide a more objective measure of self-concept. They argue that "on logical grounds the self-report measure cannot be used as a direct measure of the

self-concept" (p. 494) and that ratings by external observers should replace self-ratings as the preferred measure of self-concept. In contrast, others (e.g., Crandall, 1973; Marsh, Smith, Barnes, & Butler, 1983; Shavelson et al., 1976) argue for the theoretical separation of self-concept based on a person's own self-report from inferred self-concepts which are based on the reports of others. Crandall suggested that ratings by others may be useful to validate or even to supplement self-report measures. Marsh argued that ratings by others are phenomenologically distinct from self-concept. Thus, inferred self-concepts will agree with self-reports only if the external observer knows the subject well, observes a wide range of behaviors, has viewed a range of different subjects, and makes judgments of the same specific characteristic as the subject. Shavelson et al. predict that although there may be self-other agreement for very specific self-concepts near the base of their hierarchy, the "correspondence between observer and self decreases as one moves up the self-concept hierarchy" (page 412). For most researchers, self-concept must be based on self-perceptions which can best be examined with self-report measures. Ratings by others are not equivalent to self-ratings and will correlate highly with them only in certain circumstances.

Strauger and Schoeneman (1979) reviewed studies that correlated self-reports with judgments by others. They concluded that "there is no consistent agreement between people's self-perceptions and how they are actually viewed by others" (page 549). However, in their review the content of the self-reports was quite varied, no attempt was made to determine if some external observers (e.g., teachers, parents, peers) provided more accurate assessments than others, and the distinctiveness of different components was not considered when multiple characteristics were judged. Furthermore, no distinction was made between studies that asked external observers to record their own perceptions and those in which observers made inferred self-concept ratings.

Self-Other Agreement in SDQ Research

When multiple dimensions of self-concept are represented by both self-ratings and inferred ratings, multitrait-multimethod analysis provides a powerful analytical tool for testing the construct validity of the self-concept facets. Using this method of analysis, convergent validity is inferred from substantial correlations between self-ratings and inferred ratings on matching self-concept traits. Discriminant validity inferred from the lack of correlation between nonmatching traits provides a test of the distinctiveness of self-other agreement and of the multidimensionality of the self-concept facets.

A series of multitrait-multimethod studies by Marsh, Parker, and Smith (1983), Marsh, Smith, and Barnes, (1983, 1984), and Marsh, Smith, Barnes, and Butler (1983) demonstrated significant agreement between multiple self-concepts inferred by primary school teachers and the student's own responses to the SDQ-I. This supports the convergent validity of the SDQ-I. Student-teacher agreement tended to be highest in academic areas, where the teachers could most easily make relevant observations, and lowest on Parent Relations. Support for the discriminant validity of the SDQ-I scales was also demonstrated in that student-teacher agreement on each scale was specific to that scale and could not be explained in terms of a generalized agreement that incorporated different areas. Marsh, Smith, and Barnes (1984) collected inferred self-concepts from the students' peers, and again multitrait-multimethod analyses demonstrated moderate support for the convergent and discriminant validity of the SDQ-I scales. Soares and Soares (1977, 1982) also used multitrait-multimethod analysis to demonstrate significant self-other agreement and provide evidence for the distinctiveness of different facets of self-concept. These

studies demonstrate that external observers can accurately infer self-concepts in some circumstances.

Self-other agreement reflected in different studies using the SDQ-I is summarized in Table 14, pages 98-99. Correlations are presented between the SDQ-I scales for student responses and the corresponding scales based on responses by others (either teachers or peers). This represents the square, heterotrait-heteromethod matrix from the multitrait-multimethod studies, and the correlations along the diagonal are the convergent validities (in boldface) representing student-other agreement in matching areas of self-concept. The average of the 56 convergent validities was .30 across all the SDQ-I scales (excluding the General-Self scale). Correlations were highest for the academic self-concepts and Physical Abilities and lowest for Parent Relations and Physical Appearance.

A construct validation approach to the study of self-concept emphasizes a *pattern* of correlations in which external criteria should be more highly correlated with the facets of self-concept to which they are most logically related. In multitrait-multimethod analysis, this emphasis is embodied in the comparison of each convergent validity with other correlations in the same row or column of the same square (heterotrait-heteromethod) block. Ignoring the General-Self scale in Table 14, this involves 12 comparisons for each convergent validity in each study, or a total of 896 (12 comparisons x 7 scales x 8 studies) comparisons. For 95% of the comparisons, the convergent validity was higher than the other correlations in the same square block. Most of the failures involved the Physical Appearance and Parent Relations scales where the convergence was often so low that tests of discriminant validity are dubious. These findings illustrate that self-other agreement is specific to particular areas of self-concept and provide further support for the construct validity of the multiple dimensions of self-concept and interpretations based on the SDQ-I.

One additional multitrait-multimethod study (see Marsh, Barnes, & Hocevar, 1985; Marsh & O'Neil, 1984) was conducted with university students who responded to the SDQ-III. This study, though not based on the SDQ-I, is particularly relevant to discussion of self-other agreement. In addition to self-ratings provided by the students, the person who knew each student best provided inferred self-concepts by completing the SDQ-III as if he or she were the student. Over half the students chose one of their parents as the person who knew him or her best. Separate factor analyses of both self-ratings and responses by significant others identified the 13 dimensions of self-concept which the SDQ-III is designed to measure. For each set of responses, internal consistencies of all scales were high, whereas the average correlations among the factors were close to zero. Self-other agreement was quite high (mean $r = .58$), demonstrating that significant others are able to accurately infer the multidimensional self-concepts of someone whom they know well, and supporting the validity of interpretations based on responses to the SDQ-III.

Inspection of the multitrait-multimethod matrix for this study (see Table 15, page 100,) indicates that all 13 convergent validities (in bold italics) were statistically significant and the mean of these values (.58) was substantial. In addition, for each of the 312 possible comparisons between a convergent validity and another correlation in the same row or column of the square block of coefficients (heterotrait-heteromethod), the validity coefficient was higher. Finally, for 310 of the 312 possible comparisons between a convergent validity coefficient and other correlations in the same row or column of the two triangular blocks (heterotrait-monomethod), the validity coefficient was higher. These findings provide strikingly strong support for both the convergent and divergent validity of responses by older subjects to the SDQ-III.

Table 14. Correlations Between Multiple Facets of Self-Concept and Self-Concepts Inferred by Significant Others in Eight Studies

Inferred Ratings	Self-Ratings								
	Study	Physical	Appearance	Peers	Parent	Reading	Math	School	General ^a
Physical	(1)	38	24	37	05	-05	19	27	
	(2)	41	20	25	-06	-06	11	11	
	(3)	30	22	41	03	-06	11	10	
	(4)	32	10	19	06	02	10	07	
	(5)	50	00	29	01	-06	11	19	
	(6)	48	16	20	00	-02	05	00	17
	(7)	30	23	23	13	03	09	17	08
	(8)	32	03	16	00	-01	03	03	
Appearance	(1)	20	21	32	14	19	17	28	
	(2)	26	13	18	04	05	05	13	
	(3)	21	22	41	14	09	13	14	
	(4)	05	06	08	02	05	04	05	
	(5)	24	13	30	11	15	32	39	
	(6)	17	07	09	01	-03	01	-01	12
	(7)	22	25	34	23	08	13	23	06
	(8)	04	17	11	03	-03	01	06	
Peers	(1)	25	18	42	22	04	07	20	
	(2)	18	04	26	06	01	05	10	
	(3)	17	14	45	08	08	04	10	
	(4)	01	-02	20	01	05	01	04	
	(5)	19	-01	34	18	09	25	33	
	(6)	21	03	15	01	-01	04	04	11
	(7)	16	18	22	14	10	-01	12	13
	(8)	13	12	24	04	-04	02	00	
Parents	(1)	-04	13	03	21	21	04	09	
	(2)	-02	-09	-03	13	12	04	13	
	(3)	17	15	32	23	18	11	19	
	(4)	-06	-02	06	01	13	05	11	
	(5)	00	-16	16	18	20	30	49	
	(6)	-05	-06	04	16	-02	09	11	-01
	(7)	06	17	28	27	21	07	24	16
	(8)	00	12	09	16	05	10	16	

Note: Correlations are between multiple dimensions of student self-concept and those inferred by significant others (teachers in studies 1-7, and peers in study 8). Decimal points are omitted. Correlations in boldface (convergent validities) are the student-other agreement on matching self-concepts. Correlations among student self-concept ratings and among ratings by significant others (heterotrait-monotrait correlations) are not presented but do appear in the cited references.

^aThe revised form of the SDQ-I, which includes the General-Self scale, was only used in two of the analyses.

Table 14. (continued) Correlations Between Multiple Facets of Self-Concept and Self-Concepts Inferred by Significant Others in Eight Studies

Inferred Ratings	Self-Ratings								
	Study	Physical	Appearance	Peers	Parent	Reading	Math	School	General ^a
Reading	(1)	-08	00	03	02	40	00	24	
	(2)	-07	-14	-07	12	23	10	30	
	(3)	00	08	16	14	43	00	23	
	(4)	-10	-05	01	-05	34	06	07	
	(5)	-20	-10	03	00	52	56	51	
	(6)	00	-09	-05	-01	22	10	20	15
	(7)	-04	02	11	04	51	-03	25	13
	(8)	-08	02	02	03	28	13	21	
Math	(1)	11	04	04	-03	11	30	23	
	(2)	04	-09	-03	01	00	33	29	
	(3)	02	12	22	07	20	29	31	
	(4)	-03	-06	07	-03	24	29	21	
	(5)	02	-03	18	13	33	67	63	
	(6)	05	00	01	03	07	33	23	16
	(7)	00	03	13	15	15	37	36	17
	(8)	-02	09	12	02	10	34	26	
School	(1)	04	13	09	10	31	15	34	
	(2)	01	-10	-04	03	10	20	30	
	(3)	06	10	27	22	34	21	36	
	(4)	-06	-06	04	-03	30	18	21	
	(5)	-06	00	00	12	40	61	62	
	(6)	00	-07	-01	-01	13	22	22	16
	(7)	-06	01	07	11	32	11	27	20
	(8)	-02	08	10	06	15	25	30	
General ^b		-01	00	09	-01	23	13	14	19
Mean Convergent Validity Coefficient		03	09	15	03	06	08	06	11
		38	16	29	17	37	37	33	15

Note: Correlations are between multiple dimensions of student self-concept and those inferred by significant others (teachers in studies 1-7, and peers in study 8). Decimal points are omitted. Correlations in boldface (convergent validities) are the student-other agreement on matching self-concepts. Correlations among student self-concept ratings and among ratings by significant others (heterotrait-monotrait correlations) are not presented but do appear in the cited references.

^aThe revised form of the SDQ-I, which includes the General-Self scale, was only used in two of the analyses.

Studies in the Table include:

- (1) Marsh, Parker, and Smith (1983), sample 2, (n = 180, 6th grade);
- (2) Marsh, Parker, and Smith (1983), sample 3, (n = 125, 6th grade);
- (3) Marsh, Smith, Barnes, and Butler (1983), sample 1, time 1 (n = 152, 4th grade);
- (4) Marsh, Smith, Barnes, and Butler (1983), sample 1, time 2 (n = 152, 4th grade);
- (5) Marsh, Smith, Barnes, and Butler (1983), sample 2, time 1 (n = 548, 5th & 6th grades);
- (6) Marsh, Smith, Barnes, and Butler (1983), sample 2, time 2 (n = 548, 5th & 6th grades);
- (7) Marsh, Smith, and Barnes (1984), (N = 559, 5th grade);
- (8) Marsh, Smith, and Barnes (1984), (N = 559, 5th grade).

Table 15. Correlations Between Multiple Facets of Self-Concept and Self-Concepts Inferred by Significant Others for the SDQ-III

Self-Ratings	Ratings by Self												
	Math	Verb	Academic	Prob	Physical	Appearance	Same Sex	Opposite Sex	Parent	Religion	Hnst	Emot	General
Math	95												
Verb	-03	84											
Academic	09	40	86										
Prob	08	17	30	79									
Physical	01	-14	-17	-03	96								
Appearance	07	09	25	15	09	86							
Same Sex	14	13	16	09	08	16	86						
Opposite Sex	05	24	03	05	03	11	21	90					
Parent	05	14	-02	-14	-08	-04	18	-05	91				
Religion	-01	00	-02	-06	02	-08	10	13	05	95			
Hnst	00	06	13	-01	02	01	11	27	07	-09	74		
Emot	03	12	10	05	12	08	27	28	14	01	07	91	
General	16	07	12	10	08	14	31	28	08	01	06	28	93
Others' Ratings													
Math	77	-05	03	03	-08	-06	05	-03	07	05	-09	-06	05
Verb	-07	51	31	12	-19	18	-01	06	22	04	-10	08	-06
Academic	23	15	31	09	00	07	-04	-13	-02	14	-06	-11	00
Prob	08	20	17	52	04	12	-03	-03	-07	-08	-01	18	01
Physical	04	-17	-10	-04	78	09	-01	-11	06	08	-01	18	08
Appearance	04	-03	-06	10	06	50	05	11	15	07	06	08	19
Same Sex	01	00	-12	-11	03	06	45	12	22	09	11	27	31
Opposite Sex	01	-02	20	07	12	00	12	57	08	00	-03	30	20
Parent	07	00	-07	-10	02	-08	17	08	76	14	01	15	06
Religion	05	-08	00	06	06	-13	04	-08	02	79	09	-09	01
Hnst	-07	-05	05	-01	04	00	-04	-02	10	25	44	08	-02
Emot	02	18	-11	03	11	-01	12	23	24	05	00	62	12
General	07	-08	-00	15	12	15	21	17	15	09	04	22	41
Other's Ratings													
Math	95												
Verb	00	86											
Academic	22	32	88										
Prob	13	37	-19	82									
Physical	04	00	15	15	97								
Appearance	02	14	-05	17	20	89							
Same Sex	01	11	-01	02	09	23	90						
Opposite Sex	00	02	-03	19	16	34	37	90					
Parent	11	14	10	-01	15	08	32	09	93				
Religion	05	-06	11	03	10	04	-02	08	08	95			
Hnst	-01	17	24	03	15	15	03	17	27	27	81		
Emot	-04	12	01	19	15	19	32	24	-01	08	08	93	
General	02	13	11	22	16	41	29	21	08	16	33	22	92

Note: Correlations (presented without decimal points) greater than .19 are statistically significant ($p < .05$). Internal consistency estimates (coefficient alpha) are in boldface, and convergent validities are in bold italics.

Summary and Implications

Multitrait-multimethod studies based on the SDQ-I have found significant self-other agreement between self-concept ratings by school children and self-concepts inferred by their teachers. These findings contradict the implications of the Shrauger and Schoeneman (1979) review which suggests that self-ratings and ratings by others are nearly uncorrelated. As expected, student-teacher agreement generally tends to be stronger in academic areas of self-concept and weaker in nonacademic areas, but this summary may be oversimplistic. First, student-teacher agreement for Physical Abilities is typically as high or higher than for the academic scales. Second, the pattern of correlations representing student-peer agreement (Study 8 in Table 14) is similar to that observed with student-teacher agreement in the same study (Study 7 in Table 14). In fact, self-other agreement in only two areas, Parent Relations and Physical Appearance, differs markedly from the self-other agreement on other scales. The relative lack of agreement on the Parent Relations scale is expected, since this is the area in which teachers and peers are least likely to have an adequate basis for accurately inferring self-concepts. The lack of agreement on Physical Appearance is somewhat more surprising. It is reasonable, perhaps, that standards used by teachers as the basis for inferring Physical Appearance are different from those used by students. However, even student-peer agreement on this factor in Study 8 is among the lowest of any of the scales. This suggests that students may be using idiosyncratic standards in forming their own Physical Appearance self-concepts and that these standards may not even generalize to those that they employ in making ratings about one of their classmates. Clearly there is need for further research on the basis of the formation of Physical Appearance self-concept and how it is reported.

The results of the Marsh, Barnes, and Hocevar (1985) study with the SDQ-III provide an even stronger contradiction to the implications of the Shrauger and Schoeneman review. They also provide better support for both the convergent and discriminant validity of multiple facets of self-concept than do the SDQ-I studies described earlier, or any other research known to the author. In contrast to results based on teacher and peer ratings, the significant others in the SDQ-III study, who were predominantly parents, accurately inferred self-concepts in academic and nonacademic areas, and agreement on the Parent Relations scale was particularly strong ($r = .76$). Self-other agreement on Physical Appearance observed in this study ($r = .50$) is substantial, but still below the average for all traits ($r = .58$). Perhaps, by this age, respondents are using internal standards that are more similar to those used by significant others. There are many possible explanations for the finding of such strong self-other agreement in this study: (a) subjects in this study were older and therefore knew themselves better or based their self-responses on more objective, observable criteria; (b) both subjects and significant others responded to the same well-developed instrument; (c) self-other agreement was for specific characteristics rather than for broad, ambiguous characteristics or an overall self-concept; and (d) the significant others in this study knew the subjects better and in a wider range of contexts than the observers in most research. Two of these reasons, (a) and (d), may also explain why self-other agreement in the SDQ-III study was better than in the series of SDQ-I studies.

Of the SDQ-I studies summarized in Table 14, only the two most recent analyses included the General-Self scale. Somewhat surprisingly, self-other agreement on the General-Self scale is the lowest of all the scales for self-concepts inferred by peers and among the lowest for self-concepts inferred by teachers. This may be consistent, however, with Shavelson's suggestion that self-other agreement will be smallest for facets near the apex of his hierarchy where self-concept is less clearly tied to observable behavior. Also consistent with this suggestion is the fact that self-other agreement is somewhat poorer for the General-School scale (mean $r = .33$) than for

the other two academic scales (both means = .37). These findings are further supported in that, for the SDQ-III study (Table 15), the lowest self-other agreement is also for the General-Self and General-Academic scales. Besides offering further support to the Shavelson model, these findings have important implications for other self-concept research in which a measure of general or overall self-concept is typically the only measure of self-concept that is considered. Results such as these further demonstrate that self-concept cannot be adequately understood if its multidimensionality is ignored.

The findings described in this chapter also have important implications for the study of ratings by others. However, interpreting the self-other agreement found here in terms of theory and previous research is difficult because of the various types of inferred ratings used in different studies (see Wells & Marwell, 1976). The present findings are consistent with the Shavelson model, particularly his counter-intuitive prediction that self-other agreement would be weaker for self-concepts close to the apex of his hierarchical model. However, Shavelson intentionally deemphasized the use of ratings by others in his presentation, arguing that they may not necessarily have any close correspondence with self-report measures of self-concept. Thus, self-other agreement found here does not seem particularly relevant to the Shavelson model. As described earlier, Marsh, Smith, Barnes, and Butler (1983) and Crandall (1973) each suggested the pragmatic use of ratings by others as a means of validating self-concept measures, although the suggestions were apparently not based on any specific theoretical position. The present findings clearly offer support for this application of inferred self-concepts.

Ratings by others are most relevant to theory and research based on the symbolic interactionist perspective. Shrauger and Schoeneman (1979), on the basis of an extensive review designed to test implications of the theory, concluded that there is little consistent relationship between self-ratings and ratings by others. The results of the SDQ-I research described in this chapter contradict their generalization and provide support for the symbolic interactionists, but several qualifications are necessary. First, the prediction of self-other agreement in the symbolic interactionist theory is not clear-cut and depends on other conditions. Second, the significant others in the studies described here were asked to respond as if they actually were the subject. This is the appropriate question to ask in order to determine the ability of significant others to infer self-concepts and follows from the definition of self-concept as a person's self-perceptions. Similar results may have been obtained, perhaps, if the significant others had been asked to respond according to how they felt the subject *should* have responded, but this question requires additional research. Furthermore, the Shrauger and Schoeneman review apparently confounds studies that use the two types of ratings by others although either type may be relevant when testing different hypotheses derived from the symbolic interactionist perspective (see Kinch, 1963; Shibutani, 1961; Wells & Marwell, 1976). A more complete examination of the issue would require the determination of the subject's self-perceptions, the subject's perceptions of how the significant other perceives him or her, the significant other's actual perceptions of the subject, and the significant other's inferred perceptions of the subject's self-perceptions. Finally, the symbolic interactionist perspective posits that perceptions of others *cause* self-concept, though a causal ordering in the opposite direction may also be consistent with the theory (Kinch, 1963). However, such causal orderings are problematic to test, in that any defensible test requires longitudinal data, and the author suspects that the relationship is actually reciprocal.

Further discussion of the use of ratings by significant others to test the internal/external frame of reference model (see Chapter 7) is also relevant here. As indicated earlier, the finding that inferred self-concepts by significant others agree moderately, or even substantially, with the person's self-reports does not mean that the processes underlying their formation are similar. It was hypothesized that inferred

self-concepts in reading and math were based primarily on an external comparison process, whereas self-concepts are actually based on both internal and external comparison processes. The strong support for this hypothesis clearly indicates that different processes operate in the formation of self-concept and self-concepts inferred by others. Furthermore, at least for academic self-concepts inferred by primary school teachers, not even the external comparison process operates in the same way for the external observer and the person. Preadolescent students, when forming their own academic self-concepts, apparently use their classmates as their frame of reference, whereas teachers use a broader frame of reference that is more closely related to absolute measures of academic achievement. Thus, the academic self-concept of an average-ability student may actually be above average in a low-ability school and below average in a high-ability school, but teachers are likely to infer the student's academic self-concept to be average in both settings. This distinction has practical implications for understanding self-concepts and also argues that inferred self-concepts should not be used instead of self-concepts based on self-reports.

The enhancement of self-concept is a desirable goal in many areas of research and is posited frequently as an intervening process that may lead to desirable changes in other constructs such as academic achievement. However, research presented earlier (see Chapter 5) indicates that self-concept is — and should be — relatively stable over time. Consequently, interventions typically have little effect on self-concept. This chapter describes the methodology of intervention studies and examines the intervention effects produced by two Outward Bound self-development programs as assessed by the SDQ instruments.

Methodological Issues in Intervention Studies

Well-controlled interventions have not systematically affected 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, post-group euphoria, etc.). Scheirer and Kraut (1979; also see Byrne, 1984) reviewed intervention studies that attempted to improve self-concept as a means to improving academic achievement. They found predominantly null results in that most of the interventions failed to alter either self-concept or academic achievement. However, in the few studies that did produce positive effects, systematic parental involvement, in which parents expected and supported better academic performance by their children, was employed. Wylie (1979) reviewed studies of the effects of psychotherapy and growth-producing group experiences on self-concept and also found predominantly null results. Marsh, Richards, and Barnes (1986a, 1986b) suggest two possible reasons for this lack of success (for alternative reasons, 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 goals of the program. 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 to only a few subjects.

Marsh, Richards, and Barnes (1986a, 1986b) examined methodological issues in the study of intervention effects on multidimensional self-concepts. They identified what they called a post-group euphoria effect — the good feelings that subjects have after the completion of intensive group experiences. They did not question that such an effect existed; rather, they were concerned that its existence affected measures designed to assess the effect of the intervention, particularly self-concept measures. They argued that randomly assigned control groups provide little protection against such a bias, while placebo controls, which are similar to the program yet are unrelated to its intended effects, are unlikely to exist or may not be feasible. Instead, they presented a construct validity approach to the study of intervention effects. Using this approach, they argued that specific dimensions of self-concept most relevant to the intervention should be most affected, whereas less relevant dimensions should be less affected and serve as a control for response biases. This approach was used in the studies described in the remainder of this chapter which reviews the effects of two Outward Bound courses on dimensions of self-concept.

Study of the Outward Bound Standard Course

The Outward Bound Standard Course is a 26-day residential program for 17- to 25-year-olds that includes physically and mentally demanding outdoor activities. Richards (1977) stated that the purpose of Outward Bound courses 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" (page 69). Newman (1980) examined the Outward Bound experience in terms of theories of self-concept development, causal attributions, and environmental psychology, and concluded that "From this framework the ideal Outward Bound process emerges as a therapeutic model" (page 341).

To study the effects of the Standard Course, Marsh, Richards, and Barnes (1986a, 1986b) used the construct validity approach described above. They conducted a short multiple time-series design in which 26 different groups of participants in different Outward Bound Standard Courses completed the SDQ-III on four occasions: one month before the start of the course, on the first day of the course, on the last day of the course, and 18 months after the completion of the course. The findings demonstrated that: (a) there was little systematic change in self-concepts during the control interval (from Time 1 to Time 2); (b) changes during the experimental interval (from Time 2 to Time 3) were large for those facets of self-concept judged to be most relevant to the goals of the program (i.e., nonacademic facets); (c) changes in dimensions of less relevance to the program were significantly smaller; (d) these intervention effects were similar for 26 different groups of participants; and (e) effects were stable during the 18-month follow-up period (from Time 3 to Time 4).

These findings supported results from an American Outward Bound study by Smith, Gabriel, Schott, and Padia (1975) that also used a time-series design to show that the intervention had a positive effect on self-assertion and self-esteem (see Godfrey, 1974; Richards, 1977; Shore, 1977, for reviews of other Outward Bound research).

Study of the Outward Bound Bridging Course

The Bridging Course was developed by Outward Bound for low-achieving high school males. Richards and Richards (1981) stated 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 program of remedial teaching, normal schoolwork, and experiences likely to influence personality in general and self-concept/self-esteem in particular" (page 4).

The design of the Bridging Course was influenced substantially by McClelland's (1965) achievement motivation theory and his practical suggestions on how achievement motivation can be changed (also see Newman's 1980 analysis of the Outward Bound experience). Though a detailed discussion of McClelland's research is beyond the scope of this chapter, several aspects are particularly relevant. He stressed that the first step, even before the start of the program, is to create a belief that the program will work: "In short, we were trying to make every possible use of what is sometimes regarded as an [error] in such research — namely the Hawthorne effect, the experimenter bias, etc., because we believe it to be one of the most powerful sources of change" (page 324). McClelland conceptualized motives as learned, affectively toned, associative networks arranged in a hierarchy of importance so that the problem of improving achievement motivation becomes one of moving it up in the hierarchy by making it more salient. This can be done, according to McClelland, by setting up and conceptualizing the network, tying the network to

everyday experiences, and relating it to superordinate motives and beliefs that might interfere with its operation. McClelland stressed that change is more likely if individuals commit themselves to concrete realistic goals and keep records of progress toward the goals; he suggested that at the end of a program, participants prepare a specific statement of their own goals for the future to make the practical implications of the program more concrete and to serve as a basis for subsequent evaluation of their progress. The setting of such a program, according to McClelland, should be one where the individual is removed from his or her everyday routine, isolated from the outside world, and made to feel that he or she is "warmly but honestly supported and respected by others as a person capable of guiding and directing his [or her] own future behavior" (page 329). Finally, McClelland argued that when participants who share an intensive learning experience come from the same community, they are more likely to form a reference group that, once they return to their old environment, will reinforce the changes that have occurred.

The Bridging Course is a six-week residential program in which a small group of 11 to 16 participants — primarily ninth-grade students — is removed to an isolated environment (except for one weekend when they return home to visit their parents). The learning environment emphasizes high degrees of task orientation and teacher involvement with and support of the students. 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 until the materials challenge the most able in the group. Individual student needs are diagnosed, goals and criteria are clearly articulated, individual student progress and performance are continuously assessed, and students are actively involved in the process of setting and monitoring goal attainment so as to foster a sense of self-responsibility. Structured exercises provide each student with the opportunity to identify "stoppers" — impediments to learning and achievement — and to discover how they can be overcome. At the end of the course, students review their progress and make commitments to future goals in a letter to themselves (mailed to them three months after the end of the course) and in a letter to their parents. Many of the materials are presented in the form of innovative educational games and group exercises that cater to the interests of the participants in order to maintain a high level of enthusiasm and interest and to emphasize the practical relevance of the academic skills. Some of the outdoor physical activities from the Outward Bound Standard Course are also included in the Bridging Course, but the primary focus of the latter is on an integrated approach to academic growth.

Five Bridging Courses, one a year from 1980 through 1984, were conducted at high schools located in different parts of Australia. Each year potential students for the course were identified from among the most poorly achieving males in a single school. Those who were selected appeared to be capable of achieving at a higher level, had strong parental support for their participation in the program, and exhibited no extreme behavioral problems. The selection was based on information from school records, teacher recommendations, standardized tests, and parent interviews. During the evolution of the Bridging Course, increasing levels of parental involvement were sought and were fostered by the selection of schools in which parental concern with academic performance appeared to be strong. Particularly in the last two years, the Outward Bound philosophy was explained to parents in detail — their active support was sought, they were told to expect changes in their son's academic performance and self-concept, they were given suggestions as to how they might reinforce these changes, and they were given periodic feedback about their son's progress during the course. At the end of the program, their son wrote them a letter in which he outlined his future goals, discussed the "stoppers" that might impede his progress, and indicated how they could support his progress. In this way the parents became active participants in the intervention and were better able to reinforce the transfer of expected changes in

self-concept and achievement back to their son's old environment at the end of the course.

Marsh and Richards (in press) conducted a study of the effects of the Bridging Course. The study was similar to the study by Marsh, Richards, and Barnes (1986a, 1986b) in that: (a) it looked at the effect of a course run by Outward Bound on multiple dimensions of self-concept as measured by one of the SDQ instruments; (b) a short multiple time series design was used; (c) the generalizability of effects was examined across different course offerings of the same (or a similar) program; and (d) a construct validity approach was used to assess the validity of the findings. The Bridging Course study differed in that: (a) the primary focus of the Bridging Course is on educational objectives rather than the nonacademic goals of the Standard Course; (b) subjects were 13- to 16-year-old low-achieving males rather than self-selected 17- to 25-year-olds; (c) subjects responded to the SDQ-I rather than the SDQ-III; and (d) the academic nature of the intervention made it possible to assess the intervention effects with objective achievement tests as well as with measures of multiple dimensions of self-concept. The juxtaposition of the two studies is particularly important. The earlier study predicted and found significantly more change in nonacademic than academic areas of self-concepts because nonacademic areas of self-concept were more relevant to the Standard Course. In contrast, because of the academic aims of the Bridging Course, it was predicted that there would be significantly greater change in the academic versus the nonacademic areas of self-concept.

Sample and Design

Subjects consisted of 66 high school males, nearly all ninth-grade students aged 13 to 16, who participated in one of the five Outward Bound Bridging Courses conducted annually between 1980 and 1984. Outward Bound was invited by school personnel to conduct the program. The school then identified low-achieving males who appeared to have the potential for improved academic performance on the basis of test scores, school records, and teacher interviews. A final group of 11 to 16 of these students was selected for each course, in part on the basis of the strength of their parents' commitment to the program and its goals. (The low-achieving students identified by the school who were not finally selected did not differ systematically from those who were selected in terms of school performance or academic test performance.) The average performance of participants in reading and mathematics was 3 to 4 years behind their age level, and comparisons with available norms suggested that they were low in terms of academic and nonacademic self-concepts. During the first three years of the study, students tended to be from lower and lower-middle socioeconomic classes; many were migrants. Since the program was subsidized by the government, the fees for participation in the program were modest. During the last two years of the study, students were from primarily upper-middle-class families that were upwardly mobile, and the fees were higher.

The design of the evaluation component of this study evolved during this five-year period. In 1980, the first year the course was offered, a standardized reading test (GAPADOL; McLeod, 1972) was administered at the beginning and end of the course. In 1981, the course was evaluated with standardized tests of reading and mathematics (*Moreton Mathematics Test*; Andrews, Elkin, & Cochrane, 1972) and the *Coopersmith Self-Esteem Inventory* (SEI; Coopersmith, 1967) at the beginning and end of the course. (See Chapter 14 for a discussion of the relationship between the SDQ-I and the SEI.) In 1982, 1983, and 1984, all participants completed the SDQ-I, the SEI, and the same standardized tests of mathematics and reading achievement approximately six weeks before the start of the course, on the first day of the course, and again at the end of the course. The SDQ-I was used instead of the

SDQ-II because its simplified language seemed more appropriate for this group. For purposes of this study, the three testing occasions are called Times 1, 2, and 3, even though there were no Time 1 scores for the 1980 and 1981 courses.

No randomly assigned control group or comparison group was considered in this study because: (a) the participants, due to the selection process and the prerequisite level of parental support, were sufficiently unique so that no comparable group of students existed in the same school; (b) when nonparticipating students were asked to complete the extensive battery of measures on each of three occasions during a 12-week period in 1982, the request met with resistance, hostility, and noncompliance; and (c) there was an ethical reluctance to draw attention to the low-achieving males who were not selected for the study since it may have created the appearance that they were unlikely to benefit from the program.

Intervention Effects

The Bridging Course was expected to enhance reading and mathematics achievement and the corresponding areas of academic self-concept. Although its effects on nonacademic areas of self-concept were likely to be smaller and less predictable on an a priori basis, the program was intended to affect self-concepts in nonacademic areas as well. In particular, the substantial commitment made by parents suggests that effects on the Home (SEI) and Parent Relations (SDQ-I) scales might be positive. The intensive involvement that the boys have with each other and with Outward Bound staff during the six weeks suggests that the effect on the Social (SEI) and Peer Relations (SDQ-I) scales might be significant. The nature of the outdoor activities suggests that the effects on the Physical Abilities (SDQ-I) scale also might be positive. Nevertheless, it was predicted that the intervention would affect academic areas of self-concept more than nonacademic areas.

The study used a short multiple time series; measures were administered about six weeks before the start of the course (Time 1), on the first day of the course (Time 2), and on the last day of the course (Time 3). Neither academic achievement nor self-concept is likely to change systematically in such a short period without any intervention, and Marsh, Richards, and Barnes (1986a, 1986b) found little systematic change in self-concept during the control interval (Time 1/Time 2) for the Outward Bound Standard Course. Thus, the change in the experimental interval (Time 2/Time 3) was expected to be substantial, statistically significant, significantly more positive than corresponding changes in the control interval, and significantly larger for academic versus nonacademic facets of self-concept.

The means for all measures are presented in Table 16, pages 110-111, for Times 1, 2, and 3 separately for each course and averaged across all courses. In addition, the table includes a summary of effect sizes for changes over the control interval and the experimental interval. For the 1980 course, only the reading achievement test was administered, and there was no control interval as it was used only at Times 2 and 3. The results indicated that a large change in reading levels of almost two (age-equivalent) years took place during the experimental interval. For the 1981 course the mathematics achievement test and the SEI were added, but again no Time 1 measures were collected. Gains in reading and math achievement were almost one year, and there were improvements in the Home scale and particularly the Academic scale of the SEI. These findings clearly support the effectiveness of the Outward Bound intervention, but the lack of a Time 1 measure and the SEI's apparently dubious ability to differentiate among different components of self-concept leave these substantial effects open to alternative explanations.

For the 1982, 1983, and 1984 courses, both achievement tests as well as the SEI and the SDQ-I were administered at Times 1, 2, and 3. In each year there were

Table 16. Means and Standard Deviations for Self-Concept and Achievement Scores at Time 1, Time 2, and Time 3 for Each Year of the Study

Measure	Time	1980 (n = 11)		1981 (n = 12)		1982 (n = 12)		1983 (n = 15)		1984 (n = 6)		Total (n = 43)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SDQ Scales	1	2.91	.57	2.74	.81	2.88	.56	2.84	.56	2.84	.56	2.84	.65
	2	3.05	.70	3.12	.72	2.96	.74	3.04	.71	3.04	.71	3.04	.71
	3	3.50	.78	3.72	.61	3.61	.86	3.62	.72	3.62	.72	3.62	.72
	Effect	2		3		2		3		2		3	
Reading	1	3.22	.78	3.32	1.14	3.82	1.08	3.48	1.04	3.48	1.04	3.48	1.04
	2	3.11	.56	3.93	.98	4.12	.74	3.77	.71	4.12	.71	3.77	.71
	3	3.67	.85	4.43	.42	4.24	.92	4.14	.81	4.24	.92	4.14	.81
	Effect	2		3		2		3		2		3	
Math	1	3.04	1.19	2.67	1.05	2.63	1.01	2.76	1.07	2.63	1.01	2.76	1.07
	2	3.04	.92	3.09	1.16	2.73	1.05	2.94	1.05	2.73	1.05	2.94	1.05
	3	3.65	1.05	3.98	.68	3.53	.99	3.71	.92	3.53	.99	3.71	.92
	Effect	2		3		2		3		2		3	
Physical	1	3.61	.71	3.82	.73	4.11	.51	3.87	.67	4.11	.51	3.87	.67
	2	3.63	.59	4.06	.44	4.10	.63	3.95	.59	4.10	.63	3.95	.59
	3	3.90	.62	4.15	.62	4.23	.58	4.10	.61	4.23	.58	4.10	.61
	Effect	1		2		3		1		2		3	
Peers	1	3.26	.56	2.99	.93	3.33	.77	3.19	.78	3.33	.77	3.19	.78
	2	3.38	.55	3.00	.88	3.58	.77	3.32	.75	3.58	.77	3.32	.75
	3	3.63	.47	3.24	.91	3.72	.64	3.53	.72	3.72	.64	3.53	.72
	Effect	2		3		2		3		2		3	
Parents	1	3.04	1.19	3.81	.88	3.48	.78	3.48	.78	3.48	.78	3.48	.78
	2	3.35	.76	4.00	.65	3.69	.93	3.70	.82	3.69	.93	3.70	.82
	3	3.68	.79	3.95	.77	3.84	.71	3.83	.74	3.84	.71	3.83	.74
	Effect	2		3		2		3		2		3	
Total Academic	1	3.98	.60	4.03	1.05	4.48	.50	4.19	.78	4.48	.50	4.19	.78
	2	3.95	.74	4.54	.49	4.44	.62	4.34	.65	4.44	.62	4.34	.65
	3	4.27	.61	4.66	.43	4.65	.33	4.54	.48	4.65	.33	4.54	.48
	Effect	2		3		2		3		2		3	
Total Nonacademic	1	3.06	.65	2.91	.84	3.11	.58	3.02	.69	3.11	.58	3.02	.69
	2	3.07	.60	3.38	.78	3.27	.65	3.25	.68	3.27	.65	3.25	.68
	3	3.61	.77	4.04	.45	3.79	.73	3.82	.67	3.79	.73	3.82	.67
	Effect	2		3		2		3		2		3	
Total	1	3.48	.36	3.67	.68	3.85	.47	3.68	.54	3.85	.47	3.68	.54
	2	3.58	.36	3.90	.41	3.95	.60	3.83	.50	3.95	.60	3.83	.50
	3	3.87	.37	4.00	.50	4.11	.41	4.00	.43	4.11	.41	4.00	.43
	Effect	2		3		2		3		2		3	

Note: The missing entries in 1980 and 1981 correspond to scores that were not collected in those two years. For changes during the control (T1 T2) and experimental (T2 T3) intervals that are greater than two standard errors. Effect is represented as 1 = change occurred in the control interval only, 2 = change occurred in the experimental interval only, and 3 = change occurred in both intervals. The large number of comparisons and the small sample sizes make problematic the interpretation of statistical significance of these comparisons (see Table 18).

From Marsh, H. W., & Richards, G. (in press). The Outward Bound Bridging Course for low achieving high school males: Effects of academic achievement and multidimensional self-concepts. *Australian Journal of Psychology*. Copyright 1988 by H. W. Marsh and G. Richards. Reprinted by permission of the authors.

Table 16. (continued) Means and Standard Deviations for Self-Concept and Achievement Scores at Time 1, Time 2, and Time 3 for Each Year of the Study

Measure	Time	1980 (n = 1)		1981 (n = 12)		1982 (n = 12)		1983 (n = 15)		1984 (n = 6)		Total (n = 43)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SEI Scales	1			3.42	1.68	4.08	1.62	2.33	1.67	2.94	1.77	3.05	1.80
	2			5.83	1.74	3.75	1.82	3.73	1.39	4.63	1.99	3.93	1.75
	3			2		3.75	1.35	5.60	2.02	5.56	1.46	5.24	1.82
	Effect			2		3		3		3		3	
Social	1			5.50	1.93	5.50	1.93	5.60	2.16	5.63	1.75	5.59	1.91
	2			5.67	1.23	5.50	1.94	6.40	1.50	5.63	1.78	5.82	1.63
	3			6.08	1.68	6.75	1.77	6.80	1.97	6.25	1.61	6.47	1.74
	Effect			2		2		1		2		2	
Home	1			5.00	2.22	5.33	2.23	5.53	2.39	6.69	.87	5.91	1.96
	2			6.17	1.85	4.75	1.60	6.87	1.41	5.94	1.53	5.72	1.84
	3			2		5.42	2.07	7.47	.83	7.00	1.21	6.60	1.66
	Effect			2		2		1		2		2	
General	1			15.75	4.31	18.50	3.42	17.20	5.51	17.69	4.60	17.74	4.58
	2			16.42	3.53	19.00	2.53	19.47	4.45	18.19	3.23	18.18	3.87
	3			2		19.33	3.77	22.00	3.51	21.94	2.65	20.18	3.96
	Effect			2		2		2		2		3	
Total	1			29.83	7.60	33.42	5.79	30.67	10.12	32.94	7.36	32.28	7.98
	2			34.50	6.68	33.00	3.62	36.47	6.53	34.38	6.13	33.65	6.45
	3			2		35.25	5.32	41.87	6.86	41.00	5.09	38.56	6.72
	Effect			2		3		3		2		3	
Achievement Tests	1			11.36	2.47	11.36	2.47	12.38	1.88	14.31	1.81	12.81	2.33
	2			10.08	2.25	11.96	2.53	14.32	1.80	14.86	2.05	12.86	2.86
	3			11.98	2.58	12.99	2.50	15.67	1.45	16.19	1.12	14.15	2.77
	Effect			2		2		3		3		3	
Math	1			11.06	1.02	11.06	1.02	10.96	.83	11.69	1.01	11.26	.99
	2			10.90	1.00	10.90	1.00	11.46	.69	11.79	.99	11.13	1.13
	3			10.99	1.19	11.43	1.52	12.51	.84	12.28	.97	11.87	1.25
	Effect			2		2		3		3		2	

Note: The missing entries in 1980 and 1981 correspond to scores that were not collected in those two years. For changes during the control (T1 T2) and experimental (T2 T3) intervals that are greater than two standard errors. Effect is represented as 1 = change occurred in the control interval only, 2 = change occurred in the experimental interval only, and 3 = change occurred in both intervals. The large number of comparisons and the small sample sizes make problematic the interpretation of statistical significance of these comparisons (see Table 18).

From Marsh, H. W., & Richards, G. (in press). The Outward Bound Bridging Course for low achieving high school males: Effects of academic achievement and multidimensional self-concepts. *Australian Journal of Psychology*. Copyright 1988 by H. W. Marsh and G. Richards. Reprinted by permission of the authors.

substantial gains in both achievement tests during the experimental interval that amounted to between .5 and 1.25 years in performance. Changes over this experimental interval were also consistently substantial on the SDQ-I academic scales and the Total Academic score, whereas changes on the nonacademic scales and the Total Nonacademic score were smaller. Changes during the experimental interval for the SEI were substantial for at least one scale each year, but none of the changes on the scale or their total was consistently large for all three years.

For the 1982, 1983, and 1984 courses, unlike the earlier courses, there was also a control interval. For 1982, there were no substantial changes on any of the self-concept or achievement scores during the control interval. In 1984, there were significant changes during the control interval for reading achievement and the Academic (SEI) scale, but not on any of the other measures. However, in 1983 there were significant changes in both achievement measures, all the academic scales from both self-concept instruments, and the Home/Parent Relations scales from both instruments. The significant shifts during the control interval complicate the interpretation of the results, but they are consistent with the intention to create a placebo-like effect before the start of the program as suggested by McClelland (1965).

It was also predicted that shifts during the experimental interval would be significantly larger than shifts in the control interval and that the difference would be larger in academic areas (though significant shifts in the academic measures observed in the control interval would complicate this comparison). To test this prediction a series of repeated measures ANOVAs were conducted with the commercially available MANOVA routine from SPSS (Hull & Nie, 1981) on the 1982, 1983, and 1984 scores. For each student, shifts (i.e., difference scores) during the control interval were compared with those for the experimental interval for six sets of variables: all seven SDQ-I scales, the two SDQ-I total scores, the three SDQ-I academic scales, the four SDQ-I nonacademic scales, the four SEI scales, and the two achievement tests (see Table 17).

The first analysis consisted of a 2 (control vs. experimental interval) x 7 (SDQ-I scales) x 3 (1982, 1983, and 1984) ANOVA where the first two factors were within-subject or repeated measures factors. In support of the predictions, the effect of interval was statistically significant, but it varied with the self-concept scale. Subsequent analyses were conducted to further examine these findings. Analyses of the two SDQ-I total scores demonstrated that the interval effect was larger for the academic than for the nonacademic component. Analysis of the four SDQ-I nonacademic scales indicated no significant difference in shifts during the control and experimental intervals for any of these scales. Analysis of the three SDQ-I academic scales indicated a large interval effect that varied somewhat with the area of academic self-concept. Inspection of Table 16 indicates that the positive shift in Reading self-concept was somewhat larger than that for Math and General-School self-concepts during the control interval and smaller during the experimental interval. In a similar analysis of the four SEI scales, the effect of interval was statistically significant but did not vary significantly for the different scales, thus supporting the earlier finding that the different SEI scales are not very distinguishable. Since the interval x year and interval x scale x year interactions are not statistically significant for any of these analyses, the results are reasonably consistent across the three years in which the short time series was employed. These findings provide strong support for the self-concept predictions based on the SDQ-I, some support for predictions based on the SEI, and support for the generalizability of the findings.

In an analysis of the two achievement tests (Table 17), the effect of interval was statistically significant and did not vary with the year of the study. The lack of interval x scale interaction suggests that the difference between control and

Table 17. Repeated Measures ANOVAS Comparing Control Interval Changes with Experimental Interval Changes for Six Sets of Variables

Source	7 SDQ-I Scales			2 SDQ-I Totals			3 SDQ-I Academic Scales			4 SDQ-I Nonacademic Scales			4 SEI Scales			2 Achievement Test		
	df	MS	F ratio	df	MS	F ratio	df	MS	F ratio	df	MS	F ratio	df	MS	F ratio	df	MS	F ratio
Between Subjects																		
Year (Y)	2	1.10	1.34	2	.40	1.60	3	.39	2.80	2	.04	0.08	2	36.93	3.67*	2	8.31	8.29**
Error	40	.82		40	.25		40	.77		40	.46		40	10.06		40	1.00	
Within Subjects																		
Scale (S)	6	1.54	7.17**	1	2.40	20.91**	2	.38	2.48	3	1.16	5.40**	3	32.31	10.97**	1	22.17	40.57**
Y x S	12	.40	1.87*	2	.35	3.04	4	1.17	.17	6	.24	1.09	6	8.37	2.84*	2	.36	.47
Error	240	.21		40	.11		80	.16		120	.22		120	2.95		40	.55	
Interval(I)	1	5.66	6.00*	1	2.03	7.14**	1	9.39	9.85**	1	1.71	2.39	1	32.05	4.63*	1	5.62	6.70*
Y x S	2	.67	.71	2	.19	.52	2	.28	.29	2	.03	0.04	2	13.84	2.00	2	1.75	2.08
Error	40	.94		40	.28		40	.95		40	.71		40	6.92		40	.84	
I x S	6	1.26	3.17**	1	1.16	5.35**	2	1.57	4.19*	3	.82	2.14	3	6.24	1.03	1	1.18	1.09
Y x I x S	12	.46	1.15	2	.00	.01	4	.65	1.73	6	.29	.76	6	11.06	1.54	2	2.26	2.08
Error	240	.40		40	.21		80	.37		120	.38		120	6.07		40	1.09	

Note: The results were performed on six sets of scores. The two levels of the interval factor are the control interval and the experimental interval, and the Scale factor refers to the different scales of the self-concept instruments or to the mathematics and reading scores of the achievement tests.

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

From Marsh, H. W., & Richards, G. (in press). The Outward Bound Bridging Course for low achieving high school males: Effects of academic achievement and multidimensional self-concepts. *Australian Journal of Psychology*. Copyright 1988 by H. W. Marsh and G. Richards. Reprinted by permission of the authors.

experimental intervals was similar for mathematics and reading scores. In a final analysis (not shown), the mathematics and verbal achievement scores and the Math and Reading self-concept scores were considered together. Again, the shifts in the experimental interval were substantially larger than in the control interval ($p < .001$), but this difference did not vary significantly with the academic content area (verbal vs. mathematics), the type of measure (self-concept vs. academic achievement), the year of the study, nor any combination of these variables (all p 's $> .1$). These results provide strong support for the academic achievement predictions and suggest that the results are similar for achievement and self-concept in mathematics and verbal areas.

Discussion

The study 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 males 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, perhaps, in relation to the SEI; and (d) the effectiveness of the SDQ-I as a measure that validly reflects the effect of a powerful academic intervention. Support for the validity of responses to the SEI, particularly in relation to academic achievement, was more problematic, but self-concept as assessed by this measure also improved as a consequence of the Bridging Course.

The size and specificity of the effects observed in the control interval — primarily in 1983 and to a lesser extent in 1984 — require further consideration. As recommended by McClelland (1965), at Time 1 the project staff specifically sought to engender a belief in students and their parents that the Bridging Course would affect academic self-concept and achievement. The significant shifts during the control interval suggest that this strategy was effective. It is also important to note that both the 1983 and 1984 courses were conducted with students from the same high school where parental support was very high. The 1982 course, on the other hand, for which there was no evidence of any systematic shift in either achievement or self-concept during the control interval, was conducted for a different school where students' enthusiasm and parents' involvement were not as strong. The support of the school personnel was also stronger in 1983 and 1984. School personnel, and particularly the parents, believed that the intervention would work, and there was pressure on the Outward Bound Director to accept additional students into the program, resulting in an increase in the number of students accepted in 1983 and 1984.

Thus, both the students who were accepted and their parents were a "selected" group that was highly motivated to succeed — and very susceptible to "placebo" effects. If the study had been limited to a general measure of self-concept, the results for the control interval may have been dismissed as a "simple" placebo effect and may have undermined results from the experimental interval. However, because significant increases occurred in both areas of achievement, particularly reading, and because changes in self-concept were almost exclusively limited to the academic and the Home/Parent Relations scales, support for a different interpretation is provided. Instead of being an undesirable bias in the results, the placebo effect observed here is a valid effect as indicated by its generalizability over cognitive and affective components of academic achievement and by the fact that the initial enhancement was maintained and accelerated during the experimental phase of the study. Critelli and Neumann (1984) also take the position that the negative connotation placed on placebo effects is often undeserved and that placebos can have empirically demonstrable and desirable effects which support the aims of the intervention. Apparently, this intervention was designed to take advantage of a

placebo effect and make it work to the advantage of the program goals. From this perspective, perhaps, the placebo effect should be considered as a legitimate intervention effect rather than something that needs to be controlled in assessing the intervention effect, as was done in the comparison between control and experimental intervals. These findings also support McClelland's contention that belief that a change will occur is in itself a powerful source of change.

A host of alternative explanations exists for each of the separate effects of the study, but none appears to be plausible for explaining the total pattern of results. Changes in self-concept responses could be explained in terms of a variety of biases conceptually related to the post-group euphoria effect discussed earlier, but such an explanation does not account for the specificity of the changes in different areas of self-concept and particularly for the changes in academic achievement. The achievement effects could be explained as practice effects since the same tests were used; however, the practice effects for the achievement tests chosen were expected to be small. Also, such an explanation would probably require the Time 1/Time 2 shift to be at least as large as the Time 2/Time 3 shift and would not explain the effects on academic self-concepts. Normal growth might explain some of the achievement effects but would not explain the size of the shift, the differential shift in the experimental and control intervals, nor the shifts in academic self-concept. Regression effects might explain the direction of the achievement shifts and perhaps even the self-concept shifts; however, the size of regression effects on all variables should be small because the variables were reliable, and low-achieving students were identified on the basis of accumulated school performance rather than on any of the measures actually used in the study. Furthermore, the regression effects should be as large or larger during the control interval as during the experimental interval, and they should affect both nonacademic and academic areas of self-concept. A variety of time/location specific biases that are associated with time series designs are not viable since the effects were similar in each of the three different courses (see Cook & Campbell, 1979). Hence, a variety of internal and external threats to the validity of the interpretations of the effects do not appear to be plausible.

The results of this study also complement those reported by Marsh, Richards, and Barnes (1986a, 1986b) for the Standard Course. In both studies Outward Bound courses, albeit different types of courses, were found to enhance those self-concepts that were most specific to the aims of the respective courses and were also found to have significantly less effect on other facets of self-concept. In the earlier study the aims of the intervention were specifically nonacademic, and the changes in the academic scales served as a control against which to evaluate facets that were more relevant. In contrast, the Bridging Course Study focused on academic criteria, and it was the nonacademic facets of self-concept that served as a control for the academic facets. This study also differed in that changes in academic achievement provided an objective basis for assessing intervention effects and 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 was possible in considering either one in isolation.

This investigation is one of the few studies to find that a systematic intervention designed to enhance both academic achievement and academic self-concept was successful (see Scheirer & Kraut, 1979). It is suspected that the critical features of the study that led to its success 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 that changes would occur before the start of the intervention; (c) the strong parental support for the program and parents' expectations that the program would be successful; and (d) the use of a multidimensional self-concept scale that validly measured areas of academic self-concept which were specific to the intervention's goals and which

differentiated these from other areas of self-concept. The only other research known to the author in which an intervention had significant effects on both academic self-concept and academic achievement for adolescent students was that 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 contains randomly assigned control and placebo subjects, and the intervention was also quite different. However, the four characteristics identified above were also present in the Brookover research.

The results of this study leave unanswered the important theoretical question of the causal ordering of the self-concept and academic achievement effects. In an interpretation of their earlier research, Brookover and Erikson (1975) argue that changing the expectations and reinforcement patterns of significant others, particularly parents, will lead to a change in academic self-concept that will influence academic achievement. Though agreeing with this position, Marsh and Richards also feel that changes in academic achievement will be reflected in subsequent changes in academic self-concept, and that changes in academic self-concept which are not supported by subsequent changes in achievement will be difficult to maintain. As in the Brookover study, the intervention in this study was specifically designed to enhance both academic achievement and academic self-concepts, and the results showed that both were affected. Consistent with the design of the intervention, Marsh and Richards chose to interpret these findings as support for a model of reciprocal causal effects between academic self-concept and academic achievement such that changes in one will facilitate changes in the other. From this perspective the attempt to establish the causal priority of either academic self-concept or academic achievement may be counterproductive. To the extent that an intervention is designed to influence both academic achievement and academic self-concept and is more effective than an intervention which focuses on only one, causal predominance may not matter.

Chapter 11. Age and Sex Effects

A frequent concern of self-concept researchers is the examination of how self-concept varies with age and sex (see Wylie, 1974, 1979). In SDQ-I research interest in age and sex effects has focused primarily on theoretical implications, but the findings are also relevant for the construction of appropriate normative comparison tables. However, unless the structure of self-concept is reasonably invariant across age and/or sex groups, then age and/or sex differences in mean responses may be uninterpretable. This chapter will first consider issues of factorial invariance and then examine age and sex effects in mean self-concept responses.

Tests of Factorial Invariance

Nearly all empirical investigations of age and sex effects have considered differences in the level of self-concept — whether mean responses vary systematically according to age level and whether mean responses by males differ from those by females. However, very few studies have considered whether the factor structure of self-concept responses is invariant across age and sex. Such tests are important not only in their own right, but also because factorial invariance is an implicit assumption in the comparison of mean responses for different subgroups, so that its violation renders as dubious the interpretation of mean differences.

Dusek and Flaherty (1981) and Marsh, Barnes, Cairns, and Tidman (1984) used exploratory factor analyses to demonstrate that factor structures derived from self-concept responses were similar across age and sex. However, recent advances in the use of factor analysis indicate that exploratory factor analysis is not generally suitable for testing factorial invariance. For example, Alwin and Jackson (1981) concluded that "the issues of factorial invariance are not adequately addressed using exploratory factor analysis" (page 250) and recommended instead the use of confirmatory factor analysis (also see Marsh, 1985b; Marsh & Hocevar, 1985). The general approach of confirmatory factor analysis is to first establish the ability of the hypothesized factor structure to fit responses from each group without requiring that any of the parameter estimates (factor loadings, factor variances, etc.) be the same across groups. Then, a series of increasingly restrictive factor models is tested in which different sets of parameter estimates are required to be the same for each of the groups. To the extent that a more restrictive model — one that requires more parameter estimates to be invariant — is able to fit the data as well as a less restrictive model, support for the invariance of those parameter estimates is provided. The advantages of the use of confirmatory factor analysis over exploratory factor analysis to test for factorial invariance are well documented — the factor structure to be tested is specified by the investigator, the investigator is able to separately test the invariance of different parameters, and the ability of alternative models to fit the data may be compared. Using this approach, as described in Chapter 4, Marsh and Hocevar demonstrated that factor loadings derived from responses to the SDQ-I are relatively invariant across age.

Marsh, Smith, and Barnes (1985) also noted the advantages of confirmatory factor analysis over exploratory factor analysis but found no previous applications of

confirmatory factor analysis to test factorial invariance across self-concept responses by males and females. They proposed a series of a priori factor models (see Table 18) similar to those developed by Marsh and Hocevar. In Model 1 the pattern of parameter estimates was posited to be similar for responses by males and females, but the actual values were not assumed to be the same across samples. Inspection of the parameter estimates (not shown) and the goodness-of-fit indices (Table 18) indicated that the structure was well defined for both groups. In Model 2 factor loadings were constrained to be the same for both groups, but other parameters (e.g., factor variances and covariances) were not. The goodness-of-fit indices for this model, which is typically considered to be the minimum condition for factorial invariance, were also good, and the chi-square did not differ significantly from that of Model 1. In Models 3, 4, and 5, invariance was also tested for factor variances and covariances, error/uniqueness, and all parameter estimates (i.e., total invariance), respectively. Even the most restrictive model in which total factorial invariance was tested (Model 5) provided a good fit to the data and differed only slightly from the model in which no invariance was posited (Model 1). These results provide strong support for the contention that the SDQ-I factor structure is reasonably invariant across responses by males and females while providing a justification for the comparison of mean responses to the SDQ-I by males and females.

Table 18. Goodness of Fit Indices for the Factor Models of Factorial Invariance Across SDQ-I Responses for Male and Female Fifth Graders

Model Description	Goodness of Fit Indicators						
	χ^2	df	χ^2/df	BBI	TLI	χ^2_a	df _a
0) Null Model	11928	992	12.02	.00	.00		
1) No Invariance	1664	872	1.91	.86	.92		
2) Factor Loadings Invariant	1708	896	1.91	.86	.92	44	24
3) Factor Loadings, Factor Correlations, Factor Variances Invariant	1799	932	1.93	.85	.92	135	60
4) Factor Loadings, Error/Uniquenesses Invariant	1790	928	1.93	.85	.92	126	56
5) Total Invariance	1888	964	1.96	.85	.91	224	92

Note: The null model hypothesizes complete independence of all measured variables and provides a measure of the total covariance in the data which is used in computing the Bentler-Bonett Index (BBI) and the Tucker-Lewis Index (TLI); see Bentler & Bonett, 1980; Marsh & Hocevar, 1985, for further description of the BBI and TLI. The χ^2_a and df_a are χ^2 and df differences between Model 1 and the Model being tested (Models 2-5).

Marsh (1987b), in a methodologically oriented article, further examined the factorial invariance of responses to the SDQ-I for males and females. He randomly selected groups of 500 fifth-grade males and 500 fifth-grade females from the SDQ-I normative data. Each of these groups was randomly divided in half to form four groups called M1, M2, F1, and F2. Using procedures like those used by Marsh, Smith, and Barnes (1985), Marsh examined support for factorial invariance across groups that differed only by chance (M1 vs. M2, F1 vs. F2) and across groups that differed according to the sex of the respondents (M1 vs. F1, M2 vs. F2). Support for the factorial invariance of responses to the SDQ-I was found across all comparisons, and support was nearly as strong when both groups were of the same sex as when they were of the opposite sex.

Mean Self-Concept Responses

The purposes of this section are to review briefly previous research on age and sex effects in mean self-concept responses and to summarize previous research with the SDQ-I, including analyses based upon the entire normative sample of 3,562 students.

Research Review

Age effects in self-concept. Wylie (1979) summarized research conducted prior to 1977 and concluded that there was no convincing evidence for any age effect, either positive or negative, in overall self-concept within the age range of 6 to 50. She found that research based on the better known instruments showed virtually no evidence for age effects, and results based on idiosyncratic instruments were divided approximately equally among those showing positive, negative, and no effects. Wylie further argued that findings based on separate components of self-concept were too diverse and too infrequent to warrant any generalizations. Dusek and Flaherty (1981) also found no systematic age effects in their longitudinal study of adolescent self-concept.

However, other research since Wylie's review suggests that self-concept may decline during preadolescent and, perhaps, early adolescent years. For example, Eshel and Klein (1981) found a sharp decline in general self-concept scores with age in a cross-sectional study of self-concepts in grades 1 through 4. 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 to rank their own reading ability compared with 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 Loebel (1980) reported that self-ratings in a physical ability task — shooting a basketball — were negatively correlated with age in grades 2 through 4. Meece, Parsons, Kaczala, Goff, and Futterman (1982) reported that there is a steady decline in mathematics self-concept during junior high and high school years but that the drop for females begins sooner and is larger. Marsh, Parker, and Barnes (1985), using responses by high school students to the SDQ-II, reported that self-concepts for most of the SDQ-II scales showed a decline between grades 7 and 9 and then leveled out and increased between grades 9 and 11. In summary, despite the suggestion by Wylie, there appears to be evidence of a drop in self-concept with age during preadolescent years, although the age effect may be nonlinear during adolescent years.

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 suggested, however, that sex differences in specific components of self-concept may be lost when items are summed to obtain a total score. Dusek and Flaherty (1981), in their longitudinal study of adolescent self-concept, reported differences in specific self-concepts that were consistent with sex stereotypes: males had higher self-concepts in masculinity and achievement/leadership than females, but lower self-concepts in congeniality/sociability. Findings in the Meece et al. (1982) review suggest that females have lower math self-concepts than do males 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 tenth-grade males had consistently more positive self-attitudes (including self-concept) about mathematics than tenth-grade females, but that females had more positive self-attitudes about reading than did males. Marsh, Byrne, and Shavelson (in press) examined sex differences in senior high school students with three different academic self-concept instruments including the academic scales from the SDQ-III. For each of the three instruments, males

had significantly higher math self-concepts than females whereas females had significantly higher verbal self-concepts than did males.

Several Australian studies have found significant sex differences, but these differences apparently depend on age, the component of self-concept, and the self-concept instrument (see Marsh & Smith, 1982). Marsh, Relich, and Smith (1983) examined sex differences on the SDQ-I for fifth- and sixth-grade students in coeducational and single-sex schools and found that in both groups females had higher self-concepts in Reading and General-School and lower self-concepts in Physical Abilities, Math, and Physical Appearance.

SDQ-I Analyses

Marsh, Barnes, Cairns, and Tidman (1984). The original purpose of the Marsh, Barnes et al. (1984) study was to examine sex and age effects in self-concept. Based on research summarized above, it was hypothesized that: (a) where sex differences in self-concept occurred, they would be consistent with sex stereotypes; and (b) where age effects occurred, they would show a linear, or at least monotonic, decline with increases in age for the childhood and preadolescent range of ages considered in the study. One of the most difficult problems in this type of cross-sectional study is to demonstrate that different age groups are equivalent on all characteristics that are not specifically age related. Since this is virtually impossible to prove, responses were arranged so that any nonequivalence in age groups worked against the hypothesis of a linear age effect. This was accomplished by selecting second- and fifth-grade student responses from one set of schools, and third- and fourth-grade responses from another set of schools ($N = 658$). The youngest and oldest children in the study came from the same schools, so if these students differed systematically from the children in the other set of schools, the effect would appear to be a nonlinear age effect with self-concepts in grades 2 and 5 being systematically higher or lower.

The effects of age and sex were determined by a series of univariate and multivariate analyses. An initial multivariate ANOVA showed significant sex and age effects that depended on the self-concept factor. Consequently, separate 4 (grade levels) \times 2 (sex) ANOVAs were conducted for each of the seven SDQ-I factors and the three total scores. For each analysis (see Figure 18) at least one main effect was statistically significant, but the sex and age interaction was not significant for any of the analyses (all p 's $> .35$). Thus sex differences did not vary across the age range considered in this study. Moderate sex differences (i.e., $\text{Eta} > .20$, or 4% of the variance explained) were observed for Physical Abilities (favoring males) and Reading (favoring females), and smaller differences were observed for several other scales.

The effect of grade level was statistically significant for all but the Parent Relations scale. In all but the Peer Relations scale, the significant effects were primarily linear as demonstrated by: (a) inspection of the plots in Figure 18; (b) the finding that linear components were statistically significant while nonlinear components were not; and (c) the comparison of the total effect size ($\text{Eta} = \text{linear} + \text{nonlinear effects}$) with linear effects ($r = \text{linear effects}$, see Figure 18). For Parent Relations there was no age effect at all, and self-concepts were consistently high across the age range considered. 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. For each of the other five SDQ-I factors, and for all three total scores, self-concepts dropped consistently with increases in grade level. This decline in self-concepts was moderate in size, representing a drop of about one-third of a standard deviation between grades 2 and 5 and was similar for males and females.

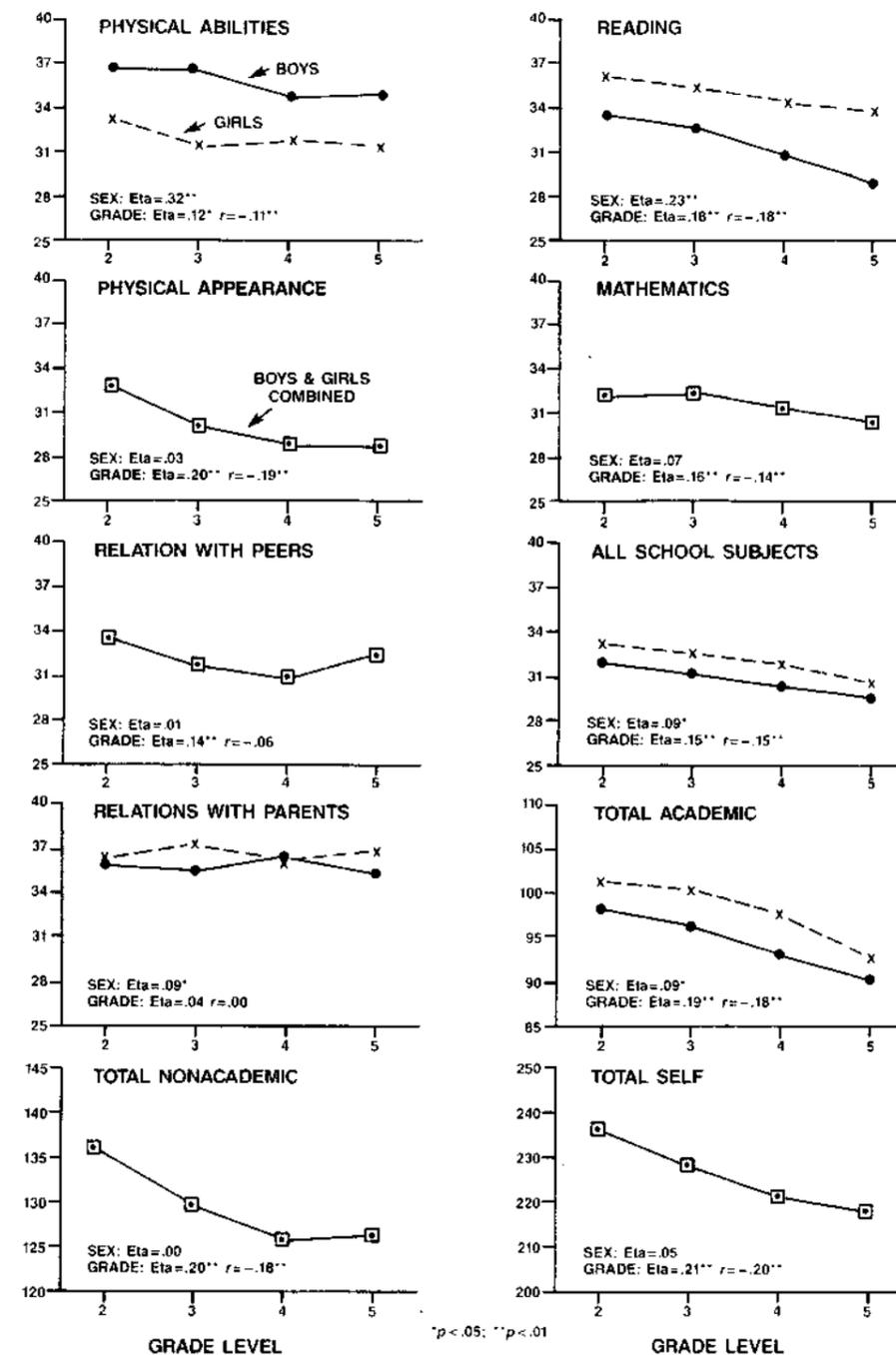


Figure 18. Grade and Sex Effects for Ten SDQ-I Scores in Grades 2-5

From Marsh, H.W., Barnes, J., Cairns, L., & Tidman, M. (1984). Self-Description Questionnaire: Age and sex effects in the structure and level of self-concept for preadolescent children. *Journal of Educational Psychology*, 76, p. 954. Copyright 1984 by the American Psychological Association. Reprinted by permission of the publisher.

Marsh, Barnes et al. (1984) argued that several characteristics of the study made the observed age effects particularly robust. First, the design of the study guarded against the possibility that age effects could be a function of nonequivalent subjects at each grade level. Second, the lack of age effect on Parent Relations — the scale with the highest level of reported self-concept in grade 2 — suggests that the age effects in other areas of self-concept were not an artifact of an age-related response bias.

The sex differences observed in this study are similar to those observed in other SDQ-I studies and are consistent with sex stereotypes suggested by other research. The lack of sex differences in the Total Self score is also consistent with the Wylie conclusion. Only the lack of sex effect in the Math self-concept score was unanticipated, but this finding may be consistent with the Meece et al. review which suggests that sex differences in Math self-concept are not well established before junior high and high school years.

Analyses of the normative data base. Analyses similar to those described above were performed on all 3,562 responses in the normative data base (see Marsh, 1985a) including the 658 responses from the Marsh, Barnes et al. study. The normative data base contains responses from students in grades 2 through 6 and thus represents a broader age range and a much larger sample. However, all the second- and third-grade students in the normative sample are the same as those in the Marsh, Barnes et al. study, and the Marsh (1985a) study had no control for nonequivalent samples as was incorporated in the earlier study.

Two sets of analyses were performed. The first was based on unweighted scale scores and the second on factor scores derived from the factor analysis of all 3,652 responses (see Table 1, page 31). The results of these analyses are summarized in Table 19. Effect sizes for main effects and interactions are summarized by eta and linear effects by the linear correlation r . The .01 significance level was used to assess statistical significance.

The results based on unweighted scale scores and factor scores are nearly identical. Partly because of the extremely large sample sizes, nearly every main effect is statistically significant. Nevertheless, the sex and age interaction reaches statistical significance for only two of the SDQ-I factors and none of the total scores, accounting for no more than .08% of the variance in any of the SDQ-I scores. This lack of interaction supports the finding reported in the Marsh, Barnes et al. study. Also, in support of the previous study, the effect of age is negative and primarily linear for all areas of self-concept and the three total scores. Here, however, the effect of age on the Parent Relations scale is statistically significant, but the effect is the smallest of any of the effects and accounts for only about 1% of the variance.

The sex effects based on the entire normative sample differ slightly from those observed by Marsh, Barnes et al. Again, the two largest sex effects are for Physical Abilities (favoring males) and Reading (favoring females). However, three other scores (Parent Relations, General-School, and Total Academic) on which females scored significantly higher, albeit only slightly, in the earlier study showed no significant sex effect in the present study. Also, four other scores (Physical Appearance, Peer Relations, Math, and Total Self) which showed no significant sex effect in the earlier study showed small effects in favor of males in the present study. Thus, it appears that females fare less well in the entire normative sample. The sizes of the sex effects, both here and in the original analysis, are generally very small. For example, for the Total Self score that represents the sum of all items (except those from the General-Self scale), the sex effect accounted for one-quarter of 1% of the variance in the original analysis and two-thirds of 1% in this analysis. The only sex effect in Physical Abilities accounted for no more than 3% of the variance.

Table 19. Effects of Age and Sex on SDQ-I Factors Represented by Unweighted Scale Scores (UWS) and Factor Scores (FS) ($N = 3,562$)

Score		Sex Effect	Age Effect	Interaction	Linear Age Effect	Total % Variance Explained
Physical Abilities	UWS	.29**	.14**	ns	-.11**	10.3
	FS	.28**	.12**	ns	-.11**	8.9
Physical Appearance	UWS	.11**	.25**	.07**	-.20**	7.2
	FS	.12**	.23**	.08**	-.20**	6.6
Peer Relations	UWS	.10**	.14**	ns	-.10**	2.7
	FS	.08**	.12	ns	-.08**	1.9
Parent Relations	UWS	ns	.12**	ns	-.07**	1.4
	FS	ns	.10**	ns	-.06**	1.1
Total Nonacademic	UWS	.18**	.23**	ns	-.18**	7.8
Reading	UWS ^a	.16**	.15**	ns	-.14**	4.6
	FS	.17**	.13**	ns	-.12**	4.6
Math	UWS	.12**	.14**	ns	-.11**	3.2
	FS	.13**	.11**	ns	-.08**	2.9
General School	UWS	ns	.19**	.06*	-.17**	3.7
	FS	.06**	.18**	.07**	-.17**	3.7
Total Academic	UWS	ns	.19**	ns	-.18**	3.8
Total-Self	UWS	.08**	.25**	ns	-.21**	6.5
General-Self	UWS ^b	.12**				

Note: Sex and age effects are represented by Eta, while the linear effect size for the age effect is represented by the Pearson r . Quadratic and cubic components of the age effects were also tested, but with the exception of a small cubic effect in Physical Appearance, none reached statistical significance at $p < .01$.

* $p < .01$; ** $p < .001$.

^aDenotes significant sex effects in which females had higher self-concepts than males.

^b $n = 739$ (mostly fifth graders); consequently, only the effect of sex was examined for statistical significance.

From Marsh, H. W. (1985). Age and sex effects in multiple dimensions of preadolescent self-concept. *Australian Journal of Psychology*, 37, p. 200. Copyright 1985 by H. W. Marsh. Reprinted by permission of the author.

Summary and Implications

Confirmatory factor analyses demonstrated that the factor structure for SDQ-I responses is relatively invariant across preadolescent age groups (Marsh & Hocevar, 1985) and sex groups (Marsh, 1987b; Marsh, Smith, & Barnes, 1985). These results are important in their own right, provide a justification for examining age and sex differences in mean responses to the SDQ-I, and further illustrate the robustness of the SDQ-I factor structure.

Age and sex effects in mean SDQ-I responses based on the entire SDQ-I normative data base were generally consistent with those found by Marsh, Barnes, Cairns, and Tidman (1984). The importance of the multidimensionality of self-concept is particularly important in examining sex effects — the direction of the effects varies with the self-concept facet, and the effects are lost when only a total score is considered. These age and sex effects are generally consistent with intuition and theory in that the sex effects are consistent with sex stereotypes, and the age effects are consistent with previous research and theoretical perspective.

Apparently there is a systematic decline in self-concepts during preadolescent years, but this should not be seen as "bad" or unfortunate. Indeed, it appears that the very positive self-concepts of the youngest children are 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) and Stipek and Tannatt (1984) described interviews with 96 children at the start of first grade in which all claimed to be among the smartest in their class. In the Marsh (1985a) study even responses by fifth- and sixth-graders averaged about 4 on the 5-point response scale. Even if the self-concepts of the youngest children are "unrealistic," this should not be interpreted to mean that their self-concepts, or responses to the SDQ-I, are biased. On the contrary, as long as their responses accurately reflect their self-perceptions, regardless of whether these self-perceptions are judged as realistic by external standards, the interpretations based on the 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. Further research is needed to identify characteristics that validly affect self-concept, develop theoretical perspectives consistent with these effects, and explore the implications of these theoretical and empirical findings.

In one such proposal that the author developed to explain the Marsh (1985a) findings and other research (see Harter, 1983; Nicholls, 1979; Ruble et al., 1980; Stipek, 1981, 1984; Stipek & Tannatt, 1984; Werner, 1957, for related theoretical positions), it was posited that very young children are egocentric and have consistently high, less differentiated self-concepts in all areas. These self-concepts may be unrealistic and relatively independent of any external criteria. As children become older, they incorporate more external information into their self-concepts so that their self-concepts become more correlated with external criteria. For most children this implies that self-concepts will decline with age in at least some areas, and that across a broad selection of children self-concepts will decline in all areas. As children incorporate more information about their actual skills and abilities, as well as feedback from others, into the formation of their self-concepts in different areas, their self-concepts will also become more differentiated as posited in the Shavelson model and observed with SDQ-I responses. This proposal is consistent with the decline in preadolescent self-concepts with age, the increased differentiation of self-facets with age, and the finding that as children become older their self-perceptions become more highly correlated with performance, performance feedback, and other external criteria.

Although this chapter has focused on childhood and preadolescent self-concepts as measured by the SDQ-I, Marsh, Parker, and Barnes (1985) also conducted a large study of adolescent responses to the SDQ-II in grades 7 to 12. Sex effects observed in that study were similar to those observed in the other SDQ research in that the sex effect was small across all facets, and sex differences in specific facets were consistent with sex stereotypes. Females had clearly lower Math self-concepts (see Marsh, Byrne, & Shavelson, in press; Stevenson & Newman, 1986). Although there were significant age effects for most self-facets, only Parent Relations showed a predominantly linear decline over this age range after showing little or no decline during preadolescent ages. For most self-facets there was a U-shaped relation with age in which self-concepts dropped between grades 7 and 9, leveled off, and then began increasing between grades 9 and 12. An extensive longitudinal study (Bachman, 1970; Bachman, O'Malley, & Johnston, 1978) also suggests that self-concepts may continue to increase during late adolescence and early adulthood. Other research summarized in the SDQ-II and SDQ-III manuals is also consistent with the suggestions that self-concept declines during preadolescence and early-adolescence, levels off during middle adolescence, and then increases from middle adolescence through at least early adulthood.

Chapter 12. Cross-National Comparisons

Unfortunately, few cross-national or cross-cultural investigations of self-concept have been conducted. For a variety of reasons, the structure and level of self-concept measured by the same instrument may vary in such comparisons. First, the connotation of the words may differ from one group to another. Second, even if the meaning of the words is the same, children's willingness to describe themselves in either favorable or unfavorable terms may differ from country to country or from culture to culture. Third, the relationships between different facets of self-concept may differ across cultures.

Cross-national comparisons have important practical implications for the generalizability of the responses to the SDQ-I and for its use outside Australia. If the SDQ-I factor structure is not reasonably well defined for responses by students from a different country, then its use may not be justified in that country. Also, if the factor structure derived from responses by preadolescents from a different country is not reasonably invariant with the structure found for Australian preadolescents, then the relations between SDQ-I factors and other constructs found in Australian studies may not generalize to research in that country. Even when the factor structures are reasonably invariant, mean responses to the SDQ-I scales by students from different countries may differ from those by Australian students; thus, the two sets of responses may not be directly comparable, and the norms tables for the SDQ-I may not be appropriate for children from other countries. Byrne and Shavelson (1986; also see Marsh, Byrne, & Shavelson, in press) recently used three academic scales and the General-Self scale from the SDQ-III, along with other multidimensional self-concept scales, with a large sample of Canadian students. They found that the SDQ-III factor structure was well defined for these four factors, but no attempt was made to compare directly the self-concepts in their study with those from Australian studies.

Smith and Marsh (1985; Marsh & Smith, 1987) collected responses to the SDQ-I by 303 English preadolescents using a random sample of nine primary schools drawn from the urban areas of Lancashire in northwest England (see Smith & Marsh, 1985, for more details of the sample). The results of that study suggested that the factor structure underlying responses to the SDQ-I by English and Australian students was similar and that group differences in the level of self-concept were small and generally nonsignificant.

The purpose of this chapter is to summarize the more detailed analysis conducted by Marsh and Smith (1987) in which recent advances in the application of confirmatory factor analysis were applied to test differences in the structure and level of self-concept for the English and Australian preadolescents. The English sample consisted of 171 males and 132 females from 11 fourth-grade classes. These students were in their final year of primary school and had a modal age of 10 years. Based on age, this school grade corresponds most closely to fifth-grade students from the SDQ-I normative sample. A random sample of fifth-grade students, also 171 males and 132 females, was selected from the normative data base to compare to the English sample. Two sets of analyses were conducted. First, the commercially available LISREL program was used to test the factorial invariance of responses to

Table 20. Goodness of Fit Indices for the Factor Models of Factorial Invariance Across Responses by Australian and English Preadolescents

Model Description	Goodness of Fit Indicators				
	χ^2	df	χ^2/df	TLI	χ^2/d_f
0) Null Model	12507.2	992	12.60	.00	
Australian	6535.6	496	13.10	.00	
English	5971.6	496	12.00	.00	
1) No Invariance	1765.6	872	2.02	.91	
Australian	889.4	436	2.04	.91	
English	876.2	436	2.01	.91	
2) Factor Loadings Invariant	1797.3	896	2.01	.91	31.7
3) Factor Loadings, Factor Correlations Invariant	1837.1	924	1.99	.85	71.5
4) Factor Loadings, Factor Correlations, Factor Variances Invariant	1846.3	932	1.98	.85	80.7
5) Factor Loadings, Error Uniquenesses Invariant	1859.8	928	2.00	.85	94.2
6) Total Invariance	1905.6	964	1.98	.85	140.0

Note: The null model hypothesizes complete independence of all measured variables and provides a measure of the total covariance in the data which is used in computing the Bentler-Bonett Index (BBI) and the Tucker-Lewis Index (TLI; see Bentler & Bonnett, 1980; Marsh & Hocevar, 1985, for further description of the BBI and TLI). The χ^2 and d_f are χ^2 and df differences between Model 1 and the Model being tested (Models 2-6).

From Marsh, H. W., & Smith, J. D. (1987). A cross-national study of the structure and level of multidimensional self-concepts: An application of confirmatory factor analysis. *Australian Journal of Psychology*, 39, p. 70. Copyright 1987 by H. W. Marsh and J. D. Smith. Reprinted by permission of the authors.

the SDQ-I by the two groups. Second, the LISREL program was used to test for group differences, sex differences, and sex x group interactions in the set of 606 responses by both groups.

Tests of Factorial Invariance

A description of the technical details of the application of confirmatory factor analysis and tests of factorial invariance is beyond the scope of this chapter, but the general procedures were similar to those described by Marsh and Hocevar (1985) in their analysis of the factorial invariance of responses to the SDQ-I by children of different grade levels. (See Chapter 11 for a brief explanation of the technique.)

All analyses described in this section were performed on covariance matrices derived from responses to the 32 item pairs that are designed to measure the eight SDQ-I factors (see Chapter 4 for a description of the use of item pairs). A restrictive "simple structure" was specified in which each item pair was allowed to load on only the factor it was designed to measure, and all other factor loadings were required to be zero. The goodness of fit indices for this model (see Table 20, Model 1) are reasonable and nearly the same for the English and Australian samples. For both samples the factor structure is well defined in that every factor loading is statistically significant and each factor accounts for a statistically significant portion of the variance (all t values ≥ 5 , $p < .001$). These results indicate that the same pattern of factor loadings is able to fit the responses by English and Australian students.

In Models 2, 3, and 4 the invariance of factor loadings, factor correlations, and factor variances, respectively, were tested (see Table 20). Goodness of fit indices demonstrate that each of the models is able to fit the data nearly as well as Model 1 (with no invariance constraints), and the chi-square differences are not statistically significant. Tests of the invariance of the factor loadings and error/uniquenesses (Model 5) and the invariance of all parameters (Model 6) also provide reasonable fits that differ little from Model 1. These findings demonstrate that the SDQ-I factor structure is very stable across responses by subjects from two different countries.

In summary, these results provide remarkably strong support for the invariance of the factor structure across the two groups and also provide a justification for making mean comparisons of self-concepts for the two groups.

Tests of Group and Sex Differences

For purposes of this analysis, three dichotomous variables representing sex (1 = male, 2 = female), group (1 = Australia, 2 = England), and the sex x group interaction were added to the 32 item pairs from the SDQ-I. Because the design is balanced (i.e., the number of males and females is equal in each group), it was possible to construct uncorrelated variables to represent each of these effects (see Cohen & Cohen, 1975). A covariance matrix representing these 35 variables was constructed from the responses by all 606 students. All analyses were conducted with an 11-factor model consisting of the eight SDQ-I factors and single-item factors representing each of the terms to be tested (see Table 21, pages 130-131). The set of 24 covariances between the eight SDQ-I factors and the three effects provides estimates of the statistical significance and direction of each relation.

A series of a priori models was tested to determine the multivariate effects of sex, group, and their interaction across the eight SDQ-I factors (see Table 22, page 132).

In Model 7 all the effects were freely estimated, and in subsequent models various combinations of effects were fixed to be zero. To the extent that a model in which effects are fixed to be zero does not differ significantly from Model 7, these effects do not have a statistically significant effect on self-concept factors. In Model 8 all eight interaction effects were fixed to be zero, and the chi-square for Model 8 did not differ significantly from that of Model 7. In Models 9 and 10 the effects of sex or group were also fixed to be zero. Whereas the test of no group differences (Model 10) was barely rejected, the test of no sex differences (Model 9) was strongly rejected. Due primarily to the significant sex effects, setting all 24 effects to be zero (Model 11) was also rejected.

One a posteriori model (Model 12), based on the results of Models 8 through 11, was specified in which seven of eight group effects, two of eight sex effects, and all eight interaction effects were fixed to be zero. The chi-square for this model was not significantly different from that of Model 1. The parameter estimates for Model 12 (see Table 21) indicate that males had higher self-concepts in Physical Abilities, Physical Appearance, Peer Relations, and General-Self and lower self-concepts in Reading and General-School. Males did not differ from females in Parent Relations and Math. Australians had higher self-concepts than did the English students in General-School, but the groups did not differ in any other areas of self-concept. There were no statistically significant group x sex interactions. However, it should be noted that the size of all these correlations is small (standardized factor covariances for the six sex effects and one group effect equal -.19, -.28, -.11, .13, .16, -.19, and -.10, respectively) even though these coefficients have been corrected for unreliability in SDQ-I responses. In particular, the correlation representing the one statistically significant group effect is only -.10.

In summary, the results of this analysis indicate that the self-concepts for samples of Australian and English students are nearly the same for all eight SDQ-I factors. Although sex effects were found for a majority of the SDQ-I factors, these effects were similar for the English and Australian samples and similar to those reported in other SDQ-I studies.

Summary and Implications

The analyses summarized in this chapter compared responses to the SDQ-I by groups of Australian and English students. In the first set of analyses, the factor structure for the two groups was found to be nearly invariant. In the second set of analyses, Australian and English students were shown to have similar self-concepts for seven of eight SDQ-I factors and to differ only modestly on the eighth. Sex differences in SDQ-I factors were found to be similar for both English and Australian preadolescents. These results provide strong support for the generalizability of the SDQ-I responses across English and Australian samples.

Several limitations in these results need to be discussed. First, although the two samples may be representative of the geographical areas from which they were chosen (i.e., urban areas of northwest England and of New South Wales, Australia), they may not be representative of children in their respective countries as a whole. Second, although the two samples represent groups in different countries, there are many cultural similarities between Australia and England. Further research is needed to determine if these results generalize to other English-speaking Western countries and to other countries where language and cultural differences are more substantial. Third, the similarity in self-concept responses by English and Australian preadolescents does not necessarily mean that their self-concepts are the same. It is possible that different processes are used to formulate self-concepts for the two groups of students but that these processes result in similar factor structures and

group average scores. As in all fields of research, interpretation of support for the null hypothesis must be made cautiously.

Smith and Marsh (1985) also discussed the suitability of the SDQ-I for English students. The SDQ-I was submitted for comment and inspection to counselors and researchers at the University of Lancaster and to the Lancashire Local Education Authority. These professionals expressed only minor reservations about a few isolated words such as "kid" and "dumb." The children themselves apparently had no trouble understanding these or any of the other words on the SDQ-I. There were no questions asked when the individual items were read aloud to students in each class, nor were difficulties raised in subsequent individual or class discussions of the instrument. These anecdotal results provide further corroboration of the suitability of the SDQ-I for English children.

Table 21. Model 12 Parameter Estimates for Tests of Sex, Group, and Sex x Group Interaction Effects for Australian and English School Children

Variable	Factor										% Specific Variance Plus Error	
	Physical	Appearance	Peers	Parents	Reading	Math	School	General	Sex ^a	Group ^a		Interaction
Factor Loadings												
Physical 1	1.00	0	0	0	0	0	0	0	0	0	0	.37*
Physical 2	.98*	0	0	0	0	0	0	0	0	0	0	.40*
Physical 3	1.01*	0	0	0	0	0	0	0	0	0	0	.35*
Physical 4	.98*	0	0	0	0	0	0	0	0	0	0	.40*
Appearance 1	0	1.00	0	0	0	0	0	0	0	0	0	.47*
Appearance 2	0	.87*	0	0	0	0	0	0	0	0	0	.60*
Appearance 3	0	1.08*	0	0	0	0	0	0	0	0	0	.38*
Appearance 4	0	1.15*	0	0	0	0	0	0	0	0	0	.30*
Peers 1	0	0	1.00	0	0	0	0	0	0	0	0	.54*
Peers 2	0	0	1.16*	0	0	0	0	0	0	0	0	.38*
Peers 3	0	0	1.80*	0	0	0	0	0	0	0	0	.36*
Peers 4	0	0	1.19*	0	0	0	0	0	0	0	0	.35*
Parent 1	0	0	0	1.00	0	0	0	0	0	0	0	.66*
Parent 2	0	0	0	.98*	0	0	0	0	0	0	0	.67*
Parent 3	0	0	0	1.31*	0	0	0	0	0	0	0	.41*
Parent 4	0	0	0	1.36*	0	0	0	0	0	0	0	.37*
Reading 1	0	0	0	0	1.00	0	0	0	0	0	0	.34*
Reading 2	0	0	0	0	1.08*	0	0	0	0	0	0	.22*
Reading 3	0	0	0	0	1.03*	0	0	0	0	0	0	.30*
Reading 4	0	0	0	0	1.05*	0	0	0	0	0	0	.28*
Math 1	0	0	0	0	0	1.00	0	0	0	0	0	.36*
Math 2	0	0	0	0	0	1.08*	0	0	0	0	0	.25*
Math 3	0	0	0	0	0	1.14*	0	0	0	0	0	.17*
Math 4	0	0	0	0	0	1.15*	0	0	0	0	0	.15*
School 1	0	0	0	0	0	0	1.00	0	0	0	0	.47*
School 2	0	0	0	0	0	0	.95*	0	0	0	0	.52*
School 3	0	0	0	0	0	0	1.08*	0	0	0	0	.38*
School 4	0	0	0	0	0	0	1.17*	0	0	0	0	.27*

Note. Parameters with values of 0 and 1 were fixed and not estimated as part of the analysis, so no tests of statistical significance were performed for these values. The χ^2 for this model did not differ significantly from the χ^2 of the model in which all sex, group, and interaction effects on self-concept were estimated (see Table 23).

^a $p < .05$.

^bSex was coded so that positive covariances indicate higher self-concepts for females. Group was coded so that positive covariances indicate higher self-concepts by English students. The standardized covariances (i.e., correlations) representing the six sex effects and one group effect are -.19, -.28, -.11, .13, .16, -.19, and -.10, respectively.

From Marsh, H. W., & Smith, I. D. (1987). A cross-national study of the structure and level of multidimensional self-concepts: An application of confirmatory factor analysis. *Australian Journal of Psychology*, 39, p. 70-71. Copyright 1987 by H. W. Marsh and I. D. Smith. Reprinted by permission of the authors.

Table 21. (continued) Model 12 Parameter Estimates for Tests of Sex, Group, and Sex x Group Interaction Effects for Australian and English School Children

Variable	Factor										% Specific Variance Plus Error	
	Physical	Appearance	Peers	Parents	Reading	Math	School	General	Sex ^a	Group ^a		Interaction
Factor Loadings												
General 1	0	0	0	0	0	0	0	1.00	0	0	0	.63*
General 2	0	0	0	0	0	0	0	1.20*	0	0	0	.46*
General 3	0	0	0	0	0	0	0	1.17*	0	0	0	.49*
General 4	0	0	0	0	0	0	0	1.21*	0	0	0	.46*
Sex	0	0	0	0	0	0	0	0	1.0	0	0	.00
Group	0	0	0	0	0	0	0	0	0	1.0	0	.00
Interaction	0	0	0	0	0	0	0	0	0	0	1.0	.00
Factor Variances and Covariances												
Physical	.64*											
Appearance	.30*	.53*										
Peers	.37*	.28*	.46*									
Parent	.13*	.12*	.18*	.34*								
Reading	.06*	.02	.10*	.08*	.66*							
Math	.13*	.15*	.14*	.06	.13*	.64*						
School	.14*	.13*	.16*	.12*	.28*	.44*	.53*					
General	.35*	.26*	.31*	.18*	.11*	.16*	.19*	.38*				
Sex	-.15*	-.21*	-.08*	0	.11*	0	.12*	-.12*	1.00			
Group	0	0	0	0	0	0	-.08*	0	0	1.00		
Interaction	0	0	0	0	0	0	0	0	0	0	1.00	

Note. Parameters with values of 0 and 1 were fixed and not estimated as part of the analysis, so no tests of statistical significance were performed for these values. The χ^2 for this model did not differ significantly from the χ^2 of the model in which all sex, group, and interaction effects on self-concept were estimated (see Table 23).

^a $p < .05$.

^bSex was coded so that positive covariances indicate higher self-concepts for females. Group was coded so that positive covariances indicate higher self-concepts by English students. The standardized covariances (i.e., correlations) representing the six sex effects and one group effect are -.19, -.28, -.11, .13, .16, -.19, and -.10, respectively.

From Marsh, H. W., & Smith, I. D. (1987). A cross-national study of the structure and level of multidimensional self-concepts: An application of confirmatory factor analysis. *Australian Journal of Psychology*, 39, p. 70-71. Copyright 1987 by H. W. Marsh and I. D. Smith. Reprinted by permission of the authors.

Table 22. Goodness of Fit Indices for the Confirmatory Factor Analysis Models Describing Tests of Sex, Group, and Sex x Group Interaction Effects for Australian and English School Children

Model Description	χ^2	df	χ^2/df	BBI	TLI	χ^2_d	df _d
0) Null Model	12019.2	595	20.20	.00	.00		
7) No Constraints	1282.4	514	2.49	.89	.92		
8) Interaction Effects Fixed to be 0	1295.8	522	2.48	.89	.92	13.4	8
9) Interaction and Sex Effects Fixed to be 0	1393.6	530	2.63	.88	.92	111.2*	16
10) Interaction and Group Effects Fixed to be 0	1316.7	530	2.48	.89	.92	34.3*	16
11) Interaction, Sex and Group Effects Fixed to be 0	1414.6	538	2.62	.88	.92	132.2*	24
12) Interaction, Selected Sex and Group Effects Fixed to be 0	1306.3	531	2.46	.89	.92	23.9	17

Note: The null model hypothesizes complete independence of all measured variables and provides a measure of the total covariance in the data which is used in computing the Bentler-Bonett Index (BBI) and the Tucker-Lewis Index (TLI; see Bentler & Bonett, 1980; Marsh & Hocevar, 1985, for further description of the BBI and TLI). The χ^2_d and df_d are χ^2 and df differences between Model 1 and the Model being tested (Models 7-12).

*p < .01.

From Marsh, H. W., & Smith, J. D. (1987). A cross-national study of the structure and level of multidimensional self-concepts: An application of confirmatory factor analysis. *Australian Journal of Psychology*, 39, p. 72. Copyright 1987 by H. W. Marsh and J. D. Smith. Reprinted by permission of the authors.

Chapter 13. The Bias of Negative Items

Responses to the 12 negative items on the SDQ-I are not included in the individual scale or total scores because research has shown that responses to these items are biased. This chapter describes research that led to this decision and discusses theoretical and practical implications of the findings. Research in this chapter is described in more detail by Marsh (1986a).

Test construction specialists argue for the use of some negatively worded items on personality, attitude, and other rating scale instruments to disrupt response biases such as responding to all items with the same response category. This is considered particularly important for single-scale instruments in which all items are designed to measure one construct. The use of negative items assumes that they measure the same construct as positive items. However, this assumption is rarely tested, and its validity seems questionable for children and preadolescent respondents. Benson and Hocevar (1985) developed three parallel scales that consisted of all positively worded items, all negatively worded items, and a mixture of the two; the positive and negative versions of the same items differed only in terms of the inclusion of the words *not* or *do not*. For preadolescent responses Benson and Hocevar found that scales defined by positive and negative items differed significantly in terms of scale means, scale variances, and scale reliabilities; subjects had difficulty responding appropriately to the negatively worded items. When both positive and negative items were included on the same form, responses were more affected by positive and negative wording than by the item content. The authors concluded that the inclusion of negative items adversely affects the validity of responses by preadolescent respondents.

To respond appropriately to negative items, respondents may have to invoke a double negative logic that requires a higher level of verbal reasoning than that required by positive items. For example, the item "I am NOT a good student" requires a response of "false" to indicate that "I am a good student." If this logic is not appropriately employed, respondents may give an answer which has exactly the opposite meaning to their intended response. However, it is important to realize that negative items do not have to be negatively worded (nor do positive items have to be affirmatively worded). For example, the affirmatively worded item "I am dumb at mathematics" is still a negative item in that a response of "true" indicates a *negative* self-concept. In the SDQ-I studies described in this chapter, a negative item bias is defined as an instance in which children respond inappropriately by indicating "true" to a negative statement when their responses to positive items have consistently indicated that the opposite response would be more appropriate, or vice versa. Such an effect will create a method/halo bias that is specific to the negative items and will bias scores based upon these items.

In the early development of the SDQ-I, unlike the SDQ-II and the SDQ-III which are designed for older subjects, negative items were found to be ineffective in defining the different areas of self-concept they were designed to measure. Preliminary analyses indicated that negative items contributed less to the internal consistency of the scales, and exploratory factor analyses sometimes revealed a negative item factor (i.e., a factor on which only negative items loaded). Younger children in particular often responded "true" to negative items, indicating a very poor

self concept, when their responses to positive items consistently indicated a positive self-concept. This suggested that the problem might be a cognitive-developmental phenomenon. In subsequent revisions considerable care was taken in the wording of the negative items so that they were clearly negative and avoided the problem of double negative reasoning as much as possible. Thus, an item such as "I do not like mathematics" was changed to "I hate mathematics." However, numerous attempts to revise the negative items failed to solve the problem and led to the recommendation that these items not be included when scoring the SDQ-I scales (Marsh, Barnes, Cairns, & Tidman, 1984).

A wide range of observations from disparate areas of research appear to be related to the negative item bias. Theoretical findings in developmental psychology and psycholinguistics may provide a basis for understanding the effect, and methodological approaches and findings from personality and achievement testing may provide research designs helpful in the study of the phenomenon. A review of the relevant research in each of these areas is beyond the scope of this chapter, but important areas are delineated by Marsh (1986a).

This chapter describes two studies (Marsh, 1986a) which demonstrate how responses to negative items are related to: (a) the SDQ-I scales as defined by positive items, (b) grade level, and (c) reading achievement. Study 1 is a reanalysis of the Marsh, Barnes et al. (1984) study in which the effect of age and sex on self-concept was examined (see Chapter 11). In the reanalysis of Study 1, the responses to negative items were added to those analyzed previously to examine the negative item bias and its relationship to age. In Study 2, a reanalysis of the study by Marsh, Smith, and Barnes (1985), tests were made of confirmatory factor analytic models in which a negative item factor was explicitly defined. Verbal ability measures were then incorporated into these models to determine how responses to the negative items are related to reading ability.

Age Effects

The purpose of the first set of analyses (Study 1) was to confirm that negative items are less internally consistent with other items in the scale they are designed to measure than are positive items. A series of item analyses were conducted for the total sample ($N = 658$) and separately for each of the four grade levels (grades 2 through 5). For the total sample the coefficient alphas for all scales and the average correlation among items within each scale were higher when the negative items were excluded. This replicated findings from earlier research. However, examination of the results for the grade-level comparison demonstrated that this effect depended upon age. The exclusion of the negative items consistently produced the largest improvement in the coefficient alphas for the youngest children. Also, the negative items formed their own scale which had reasonable internal consistency, particularly for the youngest pupils.

The total scores representing the positive and negative items, to the extent that they are measuring the same construct, should correlate about .80 or higher (i.e., within the limits of the reliabilities of the two total scores). For the total sample the correlation between responses to the two total scores was only .27, indicating that they are measuring different constructs. Furthermore, the results illustrated a dramatic developmental effect. For the youngest children, the two total scores were uncorrelated ($r = -.02$), whereas the correlation was much larger for the oldest children ($r = .60$). Thus, for the youngest children the negative items are measuring a construct that is unrelated to self-concept. For the oldest children the negative item responses are substantially related to positive item responses but still contain considerable variance that is reliable and unique. These results clearly justify the

decision to exclude the negative items in scoring the SDQ-I scales, but they also suggest that the method effect is developmentally related to the age of the subjects.

In summary, some children at each grade level seem to respond inappropriately to negative items. Across all children and preadolescents, the phenomenon is clearly age related and occurs more frequently with younger children. Because this bias is systematic rather than constant or random, it is particularly serious. These findings support the decision not to include responses from negative items in the scale scores derived from the SDQ-I, but they also have important implications for other rating scales designed for use by children and for the further study of this effect as a cognitive-developmental phenomenon.

Effects of Reading Ability

The results of Study 1 show that responses to negative items are influenced by a method/halo effect and that this effect varies with age. The negative items apparently require a higher level of verbal reasoning in order to respond appropriately, which may explain why the effect is larger for younger children. Despite the intuitive appeal of this explanation, Study 1 has important weaknesses that limit the strength of the conclusions. The use of exploratory factor analyses in the original research cannot determine whether negative items contributed to a "negative item factor," or to the scale the item was designed to measure, or to both. However, the confirmatory factor analyses described below address this issue. The suggestion that the negative item bias is systematically related to verbal reasoning or reading ability could not be tested directly since reading scores were not available. Instead, this inference was based on the finding that the negative item effect varied for different age groups and the assumption that the younger children have poorer verbal skills.

Recent advances in the application of confirmatory factor analysis provide procedures that overcome these weaknesses. In Study 2 a sample of 559 fifth-grade students completed the SDQ-I and two verbal achievement tests, while also being rated by their teachers in terms of their reading ability (see Marsh, Smith, & Barnes, 1985). Factor models were tested that required the negative items to load on the factor they were designed to measure, on a separate negative item factor, or on both. The verbal ability measures were incorporated into these models in such a way that the relationship between the negative item bias and verbal ability could be tested. Because students from only one grade level were considered, the effect of age must be minimal, and any effect of reading achievement must be relatively independent of age.

In preliminary analyses the factor structure underlying the 64 positive items from the SDQ-I (i.e., eight items from each of eight scales) was examined. For purposes of this and subsequent analyses, each scale was defined by four variables representing the total response to a pair of items, as described in Chapter 4. The results of an exploratory factor analysis of responses to the positive item pairs in this study were presented in Table 3 (page 42), and the results of a confirmatory factor analysis were also presented by Marsh, Smith, and Barnes (1985).

In the second set of analyses, responses to the 12 negatively worded items were added to the factor models. Each negative item was required to load only on the self-concept factor that it was designed to measure (Model 2.1), or only on a ninth, negative item factor (Model 2.2), or on both the self-concept factor and the negative item factor (Model 2.3). Model 2.3 provided the best fit to the data. Thus, variance in responses to the negative items represented both the factors which the items were

designed to measure and a method/halo effect represented by the negative item factor. Correlations between the negative item factor and the self-concept factors were all close to zero, and only the correlation with Reading self-concept ($r = .15$) reached statistical significance. This demonstrates that the negative item bias is statistically significant and that the method variance produced by this bias is uncorrelated with all but one of the self-concept scales.

In the third set of analyses, the three reading measures — two test scores and the teacher rating of reading ability — were added to the model. In each instance the three reading measures were used to define an additional, tenth factor called reading ability. The three reading measures were free to load on this additional factor but not on any other factor. Again, a model in which the negative items were allowed to load on both the self-concept factors and the negative item factor was able to explain the data substantially better than models in which the negative items loaded only on the self-concept factors or only on the negative item factor. For this model the parameter estimates (see Table 23, pages 138-139) for the self-concept variables — both the positive and negative items — were nearly the same as for Model 2.3 (not shown). The reading ability factor was well defined, in that each of the three variables designed to define it loaded substantially on that factor. The reading ability factor was substantially correlated with the negative item factor ($r = .42$). The reading ability factor also correlated substantially with Reading self-concept ($r = .43$) but not with any other self-concept factor.

The correlations between the reading ability factor and the other factors in Table 23 are particularly important. The negative item factor represents a method/halo bias, and these results show that this bias is substantially correlated with reading ability. Children with poorer reading skills are more likely to respond "true" to negative items than to respond in a manner consistent with their responses to positive items. The finding that reading ability is correlated with Reading self-concept, but not with other self-concept factors, further demonstrates the distinctiveness of the different self-concept factors. In summary, these findings demonstrate that negative items contribute significantly to both the scale they were designed to measure and to a negative item bias. The negative item bias is nearly uncorrelated with the self-concept factors but is substantially correlated with reading achievement.

Summary and Implications

In each study the results suggest that preadolescent children often respond inappropriately to negative items. When forced to use the more difficult reasoning required by the items, children often respond "true" or "mostly true," implying a poor self-concept, even though their responses to positive items indicate that they have favorable self-concepts. This phenomenon is more likely to occur for younger children and for children with poorer reading ability. Because most children have high self-concepts (i.e., the average response is 4 on a 5-point response scale), children who are younger and/or who have poorer reading skills will appear inappropriately to have systematically lower self-concepts than other children merely as an artifact of negative item bias. The demonstration of the substantial correlation between reading achievement and negative item bias at a single grade level indicates that the effect of reading on the bias is relatively independent of age. The negative item effect will bias interpretations of self-concept scores so that comparisons across age groups are invalid, and self-concept scores erroneously appear to be more highly correlated to reading achievement and other academic achievement scores which are frequently used to validate self-concept measures.

The focus of these studies has been on the effect of negative items as a bias to rating instruments used by children and preadolescents. However, the contention that the effect is a cognitive-developmental phenomenon was strongly supported, and further research into the substantive aspects of this effect should prove valuable. The results of Study 1 show that there is a dramatic developmental shift during early school years in the ability of preadolescent children to respond appropriately to this type of rating item. The results of Study 2 show that, within a single grade level, there were substantial individual differences in the size of the effect and that these differences were related to verbal achievement. Thus, the substantial effect of verbal achievement in Study 2 was relatively independent of age, even though the age effect in Study 1 was confounded by differences in verbal achievement. Further research is clearly needed to relate this cognitive-developmental effect to cognitive stages of early development considered in other research.

Older respondents are generally able to cope with the cognitive demands of negatively worded items, but researchers still report that positively and negatively worded items designed to measure the same construct are empirically distinct. The nature of the distinction is not clear, and there is ambiguity as to whether the distinction is a substantively important or substantively irrelevant artifact of a response bias. Marsh (1987d) more fully developed a construct validity approach to answer this question for responses to the SDQ-III. On the SDQ-III half the items for each of its 13 scales are negatively worded. Separate scores for each scale were computed for negatively worded items, positively worded items, and their total. However, differentially weighting the positive and negative components produced little or no improvement over the simple unweighted total in predicting different criteria, including inferred self-concept ratings by significant others, academic achievement scores in English and mathematics, and even retest data by the same respondents. The results provide no support for the separation of positively and negatively worded items for late-adolescent subjects but do support the contention that the negative item bias described here occurs only for children and preadolescents.

Table 23. Investigation of Negative Item Bias: LISREL Maximum Likelihood Estimates for the Self-Concept Factors, Negative Item Factor, and Reading Ability Factor

Variable	Factor								% Specific Variance Plus Error		
	Physical	Appearance	Peers	Parents	Reading	Math	School	General Self		Negative Item	Reading Ability
Factor Loadings											
Physical 1	80*	0	0	0	0	0	0	0	0	0	37*
Physical 2	85*	0	0	0	0	0	0	0	0	0	29*
Physical 3	80*	0	0	0	0	0	0	0	0	0	36*
Physical 4	80*	0	0	0	0	0	0	0	0	0	36*
Appearance 1	0	70*	0	0	0	0	0	0	0	0	52*
Appearance 2	0	63*	0	0	0	0	0	0	0	0	61*
Appearance 3	0	85*	0	0	0	0	0	0	0	0	28*
Appearance 4	0	81*	0	0	0	0	0	0	0	0	35*
Peers 1	0	0	77*	0	0	0	0	0	0	0	41*
Peers 2	0	0	79*	0	0	0	0	0	0	0	38*
Peers 3	0	0	75*	0	0	0	0	0	0	0	43*
Peers 4	0	0	82*	0	0	0	0	0	0	0	34*
Parents 1	0	0	0	48*	0	0	0	0	0	0	77*
Parents 2	0	0	0	52*	0	0	0	0	0	0	73*
Parents 3	0	0	0	80*	0	0	0	0	0	0	36*
Parents 4	0	0	0	84*	0	0	0	0	0	0	29*
Reading 1	0	0	0	0	86*	0	0	0	0	0	26*
Reading 2	0	0	0	0	88*	0	0	0	0	0	23*
Reading 3	0	0	0	0	84*	0	0	0	0	0	29*
Reading 4	0	0	0	0	84*	0	0	0	0	0	30*
Math 1	0	0	0	0	0	84*	0	0	0	0	29*
Math 2	0	0	0	0	0	88*	0	0	0	0	23*
Math 3	0	0	0	0	0	89*	0	0	0	0	21*
Math 4	0	0	0	0	0	91*	0	0	0	0	17*
School 1	0	0	0	0	0	0	78*	0	0	0	40*
School 2	0	0	0	0	0	0	64*	0	0	0	59*
School 3	0	0	0	0	0	0	80*	0	0	0	35*
School 4	0	0	0	0	0	0	84*	0	0	0	30*
General Self 1	0	0	0	0	0	0	0	66*	0	0	57*
General Self 2	0	0	0	0	0	0	0	75*	0	0	44*
General Self 3	0	0	0	0	0	0	0	79*	0	0	38*
General Self 4	0	0	0	0	0	0	0	71*	0	0	50*

Note: The three variables which define the Reading Ability factor are the Word Knowledge Test (Reading Ability 1), Reading Comprehension Test (Ability 2), and teachers' ratings of reading ability (Reading Ability 3). Parameter values of 0 and 1 were fixed and not estimated in the analysis.

*p < .01.

From Marsh, H. W. (1986). Negative item bias in rating scales for preadolescent children: A cognitive-developmental phenomena. *Developmental Psychology*, 22, pp. 46-47. Copyright 1986 by the American Psychological Association. Adapted by permission of the publisher.

Table 23. (continued) Investigation of Negative Item Bias: LISREL Maximum Likelihood Estimates for the Self-Concept Factors, Negative Item Factor, and Reading Ability Factor

Variable	Factor								% Specific Variance Plus Error		
	Physical	Appearance	Peers	Parents	Reading	Math	School	General Self		Negative Item	Reading Ability
Factor Loadings											
Negative Items 1	58*	0	0	0	0	0	0	0	35*	0	52*
Negative Items 2	0	17*	0	0	0	0	0	0	30*	0	89*
Negative Items 3	0	0	46*	0	0	0	0	0	23*	0	74*
Negative Items 4	0	0	0	16*	0	0	0	0	29*	0	88*
Negative Items 5	0	0	0	0	45*	0	0	0	34*	0	63*
Negative Items 6	0	0	0	0	63*	0	0	0	13*	0	57*
Negative Items 7	0	0	0	0	0	62*	0	0	02	0	62*
Negative Items 8	0	0	0	0	0	53*	0	0	37*	0	53*
Negative Items 9	0	0	0	0	0	0	40*	0	31*	0	72*
Negative Items 10	0	0	0	0	0	0	31*	0	53*	0	59*
Negative Items 11	0	0	0	0	0	0	0	44*	44*	0	58*
Negative Items 12	0	0	0	0	0	0	0	39*	45*	0	62*
Reading Ability 1	0	0	0	0	0	0	0	0	0	79*	37*
Reading Ability 2	0	0	0	0	0	0	0	0	0	91*	18*
Reading Ability 3	0	0	0	0	0	0	0	0	0	62*	61*
Correlations Among Factors											
Physical	41*										
Appearance	64*	48*									
Peers	41*	27*	46*								
Parents	16*	07	16*	06							
Reading	27*	21*	24*	24*	13*						
Math	39*	26*	33*	31*	43*	74*					
School	76*	50*	79*	53*	27*	40*	56*				
General-Self	06	-05	00	10	15*	12	10	08			
Negative Item	-16*	-19*	-05	-10	43*	-04	-02	-01	42		
Reading Ability											

Note: The three variables which define the Reading Ability factor are the Word Knowledge Test (Reading Ability 1), Reading Comprehension Test (Ability 2), and teachers' ratings of reading ability (Reading Ability 3). Parameter values of 0 and 1 were fixed and not estimated in the analysis.

*p < .01.

From Marsh, H. W. (1986). Negative item bias in rating scales for preadolescent children: A cognitive-developmental phenomena. *Developmental Psychology*, 22, pp. 46-47. Copyright 1986 by the American Psychological Association. Adapted by permission of the publisher.

Chapter 14.

Relationship Between Responses to the SDQ-I and Other Instruments

A useful approach to validating responses to an instrument is to correlate them with responses from other instruments designed to measure similar constructs. When both instruments are based on responses by the same person, it may be dubious to interpret convergence between the two instruments as support for the *external validity* of the different instruments. Nevertheless, tests of the consistency of responses across independently constructed instruments provide valuable information about the construct validity of responses to both instruments. Because few self-concept instruments have a well-defined factor structure, this approach is usually applied to correlations between total scores (see Wylie, 1974). A much stronger test of construct validity is possible, however, when both instruments are designed to measure similar constructs. Using the logic of multitrait-multimethod analysis, scores for the matching scales from different instruments should be substantially correlated (i.e., should have convergent validity), whereas scores between nonmatching scales should be substantially less correlated (i.e., should have divergent validity). Even when the scales from the different instruments are not strictly parallel, the logic of multitrait-multimethod analyses can be adopted (see Marsh, in press; Marsh and Smith, 1982).

Historically, researchers typically have been unable to clearly identify the factors that any one self-concept instrument is supposed to measure. Hence, little effort was made to cross-validate responses to the specific scales from different instruments. Whereas self-concept is frequently posited to be multidimensional, until recently researches have emphasized global measures of self-concept, and support for this multidimensionality was limited. Early factor analytic studies of self-concept (e.g., Coopersmith, 1967) failed to identify domain specific factors. Similarly, attempts to establish the divergent validity of domain specific measures of self-concept were typically unsuccessful.

In a classic application of the multitrait-multimethod approach to this issue, Marx and Winne (1978, also see Winne, Marx, & Taylor, 1977) related responses from three widely used self-concept instruments (the *Piers-Harris Children's Self-Concept Scale*, the *Sears Self-Concept Inventory*, and Gordon's *How I See Myself Scale*). They began by subjectively classifying scales from each instrument into one of three domains (physical, social, and academic) posited in the Shavelson et al. model. They then constructed a multitrait-multimethod matrix in which the different self-concept traits were the multiple traits and the different self-concept instruments were the multiple methods. Marx and Winne, however, found little support for the divergent validity of responses from any of the three instruments and interpreted the results as support for the unidimensionality of self-concept.

Marsh and Smith (1982) also used the multitrait-multimethod approach to examine the responses to two self-concept instruments (*Coopersmith Self-Esteem Inventory* and the *Sears Self-concept Inventory*) that were each administered on two occasions. A content analysis of the items suggested that there was only moderate overlap between the two instruments, and so it was not surprising, perhaps, to find that not even the total scores from the two instruments were highly correlated. There was little or no support for the divergent validity of responses to the two

instruments and only weak support for their convergent validity. The authors expanded the typical multitrait-multimethod approach by considering support for the convergent and discriminant validity of responses to the same instrument over two occasions. That is, different occasions served as the multiple methods, and convergence actually reflected the stability of responses to the same scale over time. Even for this apparently minimal test of construct validity, support for the discriminant validity of the Coopersmith instrument was weak though there was better support for the divergence of responses to the Sears instrument. More generally, Marsh (in press) argued that multitrait-multimethod studies that include two sources of method variance in the same study (i.e., multiple instruments and multiple occasions) are stronger than those containing only one.

In contrast to much of this earlier research, studies summarized in the three SDQ manuals provide clear support for the multidimensionality of self-concept. The purpose of material presented in this chapter is to summarize studies that have related responses from one of the SDQ instruments to responses from another self-concept instrument. Some of the most important research of this type has been conducted using the SDQ-III. This research is presented in the SDQ-III Manual and is summarized only briefly here. Two studies based on responses to the SDQ-I are presented in greater detail.

SDQ-III Research

Research summarized in the SDQ-III Manual provides support for the convergent and divergent validity of the SDQ-III responses with respect to other self-concept instruments developed for use by late-adolescents. Byrne and Shavelson (1986) and Marsh, Byrne, and Shavelson (in press) examined relations between Verbal, Math, General-School, and General-Self scales from the SDQ-III, Brookover's Self-Concept of Ability Scale (see Shavelson & Bolus, 1982), the *Affective Perception Inventory* (Soares & Soares, 1977), and Rosenberg's Self-Esteem Scales (Rosenberg, 1965). Convergent validities for all the instruments (mean $r = .68$) were consistently high, although those involving the SDQ-III were slightly higher (mean $r = .70$). Correlations between non-matching scales from different instruments (mean $r = .22$) and between different scales from the same instrument (mean $r = .25$) were substantially smaller than the convergent validities. This support for discriminant validity was also somewhat stronger for the SDQ-III responses than for responses to the other instruments. Results from each of the instruments provided strong support for the separation of Verbal and Math self-concept scales as well as the separation of academic scales and the General-Self scale (also see Chapter 7 in this Manual).

Marsh and Richards (in press) compared responses to the SDQ-III and the *Tennessee Self-Concept Scales* (TSCS; Fitts, 1965). Whereas the results provided good support for the convergent and discriminant validity for most of the scales from both instruments, further analyses revealed anomalies in several of the TSCS scales. Factor analyses of the TSCS responses suggested that some of its scales contained differentiable components. When these distinguishable subcomponents of the TSCS were considered separately, there was better support for the convergent and discriminant validity of responses to the SDQ-III and TSCS instruments. For example, the TSCS Physical scale contained items related to physical ability and physical appearance. When these components were separated, there was a much clearer relationship with responses to the SDQ-III, which contains separate Physical Abilities and Physical Appearance scales. In their evaluation of the TSCS, Marsh and Richards commented on the lack of any academic self-concept scales on the TSCS instrument, indicating that this was not reasonable for an instrument designed for use by school-aged subjects.

Taken together, these two studies provide support for the construct validity of most of the SDQ-III scales. Byrne and Shavelson (1986) and Marsh, Byrne, and Shavelson (in press) provided good support for the construct validity of the academic scales and particularly for the separation of Math and Verbal self-concepts. Marsh and Richards (1988) provided support for the construct validity of many of the nonacademic SDQ-III scales.

Relationships with the Coopersmith Self-Esteem Inventory

Marsh and Richards (in press) used responses to the SDQ-I and the *Coopersmith Self-Esteem Inventory* (SEI; Coopersmith, 1967) to evaluate the effects of an intervention designed to enhance self-concept (see Chapter 10). Students in that study were high school males between 13 and 16 years of age. They were selected as being low-achieving males who appeared to have the potential for improved academic performance.

Stability and Validity

Because of the special nature of the students in this study, it was important to establish the stability and validity of the measures. The study also provided an opportunity to compare SDQ-I responses with those of one of the most widely used self-concept instruments, the Coopersmith SEI. The 43 students in the 1982, 1983, and 1984 Outward Bound courses completed each of the measures (the SDQ-I, SEI, and achievement measures) on two occasions before the start of the intervention. The correlations among these measures are presented in the form of a multitrait-multimethod matrix (see Table 24, page 144). In multitrait-multimethod analyses the same construct is measured with two or more methods; convergence refers to agreement between the same construct assessed by different methods, and divergence refers to the distinctiveness of the multiple traits. Similar analyses have been applied in situations in which the "different" methods are really quite similar (two different testing occasions), are moderately different (two different instruments), and even refer to distinct constructs (self-concept and academic achievement). Nevertheless, the logic of multitrait-multimethod analyses and the criteria developed by Campbell and Fiske (1959) can be applied in each situation, and it is argued elsewhere (e.g., Marsh, Barnes, & Hocevar, 1985; Marsh, Smith, Barnes, & Butler, 1983) that it is better to consider more than one type of method difference in the same study. For purposes of discussion, convergence and divergence will be examined separately in relation to stability coefficients for each measure, correlations between matching traits from the two self-concept instruments, and correlations between self-concept measures and achievement test scores.

Stability of the SDQ-I scales. For purposes of this analysis the multitrait-multimethod matrix consisted of the 14 x 14 correlation matrix relating the seven SDQ-I scales administered at Time 1 to those at Time 2. The stability coefficients in Table 24 are presented in boldface. In the application of the four Campbell-Fiske criteria, the seven stability coefficients (median $r = .69$) were all substantial. Hence, there is good support for stability over this six-week interval. Each stability coefficient was higher than other correlations in the same row and column of the square submatrix relating Time 1 and Time 2 measures (median $r = .18$) for 83 of 84 comparisons, and was higher than correlations among the different SDQ-I scales at Time 1 (median $r = .23$) and at Time 2 (median $r = .23$) for 81 of 84 comparisons. These results compare favorably with the similar sort of analysis described earlier (see Table 6, page 51), and provide good evidence for the stability of SDQ-I responses during the control interval and for the distinctiveness of the SDQ-I scales.

Table 24. Correlations among SDQ-I, the Coopersmith Self-Esteem Inventory, and Achievement Scores For Time 1 and Time 2

	SDQ-I Time 1							SDQ-I Time 2							SEI Time 1				SEI Time 2				Achievement							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26				
SDQ-I Time 1																														
1 School																														
2 Reading	34																													
3 Math	72	07																												
4 Physical	03	09	-10																											
5 Appearance	24	18	-01	45																										
6 Peers	28	19	40	22	29																									
7 Parents	20	40	12	31	13	24																								
SDQ-I Time 2																														
8 School	69	27	54	12	23	36	11																							
9 Reading	19	76	03	31	06	29	53	24																						
10 Math	69	17	85	10	08	41	14	80	17																					
11 Physical	01	06	-15	73	37	12	31	10	31	01																				
12 Appearance	15	06	07	44	68	05	19	19	01	17	56																			
13 Peers	-01	04	-14	41	41	45	12	11	15	02	54	35																		
14 Parents	-06	32	-04	23	-09	15	69	-06	56	05	35	01	18																	
SEI Time 1																														
15 Academic	50	45	17	27	45	06	09	23	19	22	22	38	17	07																
16 Social	24	19	39	35	50	56	40	47	16	51	34	39	53	09	41															
17 Home	61	27	16	20	15	02	79	08	34	14	26	26	09	49	15	43														
18 General	23	45	40	22	44	42	23	36	30	38	18	30	37	05	64	70	34													
SEI Time 2																														
19 Academic	23	13	18	23	-06	19	23	24	27	19	41	15	22	16	22	40	25	32												
20 Social	07	-02	05	27	27	30	10	28	-07	19	31	31	62	-01	12	42	-03	10	01											
21 Home	13	20	12	12	-18	13	29	18	35	29	31	03	08	56	07	17	28	11	05	07										
22 General	38	03	34	28	27	52	02	42	-02	46	23	34	40	-04	49	60	01	62	22	41	29									
Achievement Time 1																														
23 Reading	22	46	03	11	-03	16	37	07	49	01	10	08	11	19	06	06	31	15	19	14	09	-05								
24 Math	46	02	46	13	08	27	21	34	07	38	09	18	13	-07	72	44	33	40	43	21	-01	27	42							
Achievement Time 2																														
25 Reading	18	33	-03	00	-07	33	23	16	43	03	12	-02	21	07	-06	04	20	11	21	17	12	06	85	40						
26 Math	35	-07	44	-04	-16	27	09	21	-01	54	00	06	13	01	-07	21	18	13	33	32	09	25	41	70	46					

Note: Decimal points are omitted. Correlations larger than .29 are statistically significant ($p < .05$). Stability coefficients are in boldface. The convergent coefficients are in bold italics, and each is predicted to be higher than any other correlation in the same row and column of the rectangular matrix in which it appears. From Marsh, H. W., & Richards, G. (in press). The Outward Bound Bridging Course for low achieving high school males: Effects of academic achievement and multidimensional self-concepts. *Australian Journal of Psychology*. Copyright 1988 by H. W. Marsh and G. Richards. Reprinted by permission of the authors.

Stability of the SEI scales. For purposes of this analysis, the multitrait-multimethod matrix consisted of the 8 x 8 correlation matrix relating the four SEI scales administered at Time 1 to those at Time 2. However, in the application of the first Campbell-Fiske criterion, only two of the four stability coefficients (median $r = .39$) were statistically significant, and only the stability of the General scale was greater than .42. This lack of support for the convergence of the scales over time makes the application of the other criteria problematic. Inspection of the correlations suggests that the General scale was the only one to consistently pass the other Campbell-Fiske criteria. These results suggest that the specific subscales of the SEI may not be sufficiently stable and distinct to be interpreted separately from the total score which reflects a general dimension of self.

Stability of the achievement scores. The stability coefficients of both the reading (.85) and mathematics (.70) achievement tests were substantial, while the four correlations between the two tests (median $r = .42$) fell between .40 and .46. Thus, while reading and mathematics achievement are moderately correlated, they are also distinguishable components of academic achievement.

SDQ-I/SEI agreement. Correlations between the seven SDQ-I and four SEI scales appear in four 7 x 4 rectangular submatrices in Table 24 (i.e., Time 1 SDQ-I scales and Time 1 SEI scales, Time 1 SDQ-I scales and Time 2 SEI scales, etc.). Since the two self-concept instruments do not measure the same traits, the Campbell-Fiske criteria cannot be applied literally. Nevertheless, several of the scales from the two instruments do seem to measure similar constructs: the SEI Academic scale and the three academic scales from the SDQ-I; the SEI Social and SDQ-I Peer Relations scales; and the SEI Home and SDQ-I Parent Relations scales. Applying the Campbell-Fiske logic, each of these correlations (presented in bold italics in Table 24) should be higher than the other correlations in the same row and column of the 7 x 4 submatrix where it appears. Convergent validities relating the SEI Home and SDQ-I Parent Relations scales (median $r = .53$) passed this test for all 28 comparisons, and convergent validities relating the SEI Social and SDQ-I Peer Relations scales (median $r = .50$) satisfied 27 of 28 comparisons. Thus, this analysis provides support for the convergent and discriminant validity of the Home/Parent and Social/Peer scales from the two instruments. There was little support for the convergence of the SEI Academic scale and the three SDQ-I academic scales, and only 2 of these 12 correlations reached statistical significance.

Self-concept/academic achievement correlations. Correlations between the achievement tests and the SEI scales generally did not reach statistical significance; only 2 of the 8 correlations relating the SEI Academic scale to achievement were significant (mean $r = .15$). The correlations between the achievement test scores and SDQ-I scales demonstrated a systematic pattern of relations. Reading achievement scores were significantly correlated with each Reading self-concept score (median $r = .43$). Similarly, mathematics achievement scores were significantly correlated with each Math self-concept score (median $r = .40$). The General-School self-concept scores were less highly correlated with the achievement test scores. Correlations between reading achievement and Math self-concepts (median $r = .03$) and between mathematics achievement and Reading self-concepts (median $r = .00$) were all close to zero. Only two of the 32 correlations relating nonacademic self-concepts to the achievement scores (median $r = .10$) reached statistical significance. Thus, academic achievement in reading and mathematics was significantly correlated with academic self-concept in the same area, less correlated with other areas of academic self-concept, and not significantly correlated with nonacademic areas of self-concept. These results are consistent with other SDQ-I research (see Table 7, pages 54-56).

Summary of stability and validity analyses. The results of research with the SDQ-I scales provide good support for their stability over time and for the

systematic pattern of relationships with academic achievement in reading and mathematics. In contrast, the results suggest that the SEI scales are not stable over time and are not significantly related to achievement in reading or mathematics. Surprisingly, support for the convergent and divergent validity of the Social and Home scales of the SEI was better when related to the SDQ-I scales than when related to two administrations of the same SEI instrument. Although this does provide some support for these two SEI scales, it is probably explicable in terms of the superior performance of the SDQ-I.

The specificity of the relations between Reading and Math self-concepts and the corresponding areas of academic achievement is consistent with earlier analyses (see Chapter 6), and the lack of correlation between Reading and Math self-concepts is consistent with the internal/external frame of reference model (see Chapter 7). However, the results are particularly compelling because the range of academic achievements for these students is so truncated and because the nature of the sample would make it likely that additional factors would complicate the pattern of relations.

Relationships with the Harter Perceived Competence Scale

Shavelson et al. (1976) posited a multifaceted, hierarchical model of self-concept that provided the theoretical basis for the development of the SDQ instruments. Harter (1982, 1983) also addressed many of these issues in her review of self-concept theory and research. For example, she argued for the need to consider both domain specific components and a general, superordinate component of self. Using a theoretical perspective similar to the Shavelson et al. model, Harter developed the Perceived Competence Scale for Children (PCS). Research with the PCS (Harter, 1982, 1983), though not formally based on the Shavelson model, provides support for many aspects of the model.

Harter (1982) focused on perceived competence and hypothesized that children do not feel equally competent in all skill domains. In seeking the critical domains for elementary school children, she chose to assess the social, physical, and cognitive domains. She further hypothesized that children (age 8 and older) "have also constructed a view of their general self-worth as a person, over and above these specific competence judgments" (p. 88) and thus included a fourth general scale on her instrument. Factor analytic results clearly supported the separation of the four scales. In separate factor analyses of responses by students in fourth through ninth grades, Harter (1982) found reasonably similar factor loadings, though factor loadings were somewhat less congruent for responses by third grade students. The PCS may not be appropriate for children less than 8 years old (Harter, 1982, 1983; Silon & Harter, 1985), and Silon and Harter found that the a priori PCS structure was not well defined for responses by educably mentally retarded children who were older than 8 but had mental ages of less than 8.

Based on her 1982 factor analytic results, Harter (1983) concluded that: "Given the repeated demonstrations of this stable factor structure, we cannot concur with Winne, Marx and Taylor (1977), who find little evidence that children make distinctions between physical, social, and academic facets of self-concept" (p. 331). She further suggested that the multitrait-multimethod study by Winne et al. failed to find support for divergent validity because there was little a priori attention given to the construction of items to adequately represent the physical, social, and academic domains on the instruments used in that study. Evidence was not yet available, however, in which responses to different, more suitable self-concept instruments did demonstrate convergent and discriminant validity with respect to

these content domains. Because both are designed for use by children and claim to measure the domain specific components considered by Winne et al., the SDQ-I and the PCS appear to be well-suited for this purpose.

Marsh and Gouvenet (in press) conducted a multitrait-multimethod study of the relationship between SDQ-I and PCS responses. In preliminary analyses, separate factor analyses were used to test the a priori factors that each instrument was designed to measure. Then, multitrait-multimethod analysis of correlations between responses to the two instruments was used to test their convergent and discriminant validity. Because both the SDQ-I and PCS are designed to measure physical, social, and academic self-concepts, the multitrait-multimethod analysis resembles the classic multitrait-multimethod studies in which Marx and Winne (1978) concluded that children were unable to distinguish between these areas of self-concepts. Also considered were verbal and mathematical achievement measures, and Ryan and Connell's (1988) measure of academic motivation (also see Connell & Ryan, 1984; Ryan, Connell, & Grolnick, in press). Using the logic of multitrait-multimethod analysis, each of these additional measures should be substantially more correlated with academic self-concept measures than with nonacademic self-concept measures.

Methods

Subjects were the 508 students (42% female; 58% male) attending grades 7, 8, or 9 at one of two high schools in metropolitan Sydney. All students — except those who were absent when the materials were administered — were included in the study. Students in both schools came from predominantly middle class families representing a wide variety of ethnic backgrounds. In both schools, classroom teachers administered the self-concept instruments using written instructions. Students' multidimensional self-concepts were measured with Harter's (1982, 1983) PCS and the SDQ-I. The PCS is designed to measure four self-concept factors (Physical, Social, General, and Cognitive) whereas the SDQ-I is designed to measure the eight self-concept factors described in Chapter 2.

On the PCS each "item" actually consists of two logically opposed statements (e.g., some kids often forget what they learn; other kids can remember things easily). The child first decides which statement is most like him or her and then indicates whether that statement is "really true of me" or "sort of true of me." Responses are scored on a 1 to 4 continuum in which 4 represents the highest level of perceived competence (i.e., the positively worded item is really true of me). The PCS consists of responses to 28 items (i.e., 56 statements) and 7 items are used to infer each of the four domain-specific scales. For purposes of factor analysis, Harter (1982) factor analyzed responses to each of the 28 items, and this procedure was used here as well. Factor analyses of SDQ-I responses were based on the 32 item-pair scores described in Chapter 4 (see Table 3, page 42).

For purposes of the multitrait-multimethod analyses, the Physical Abilities, Peer Relations, and General-Self scales from the SDQ-I were posited to correspond to the Physical, Social, and General scales from the PCS. The sum of the SDQ-I academic factors (General-School, Reading, and Mathematics) was posited to correspond to the PCS Cognitive scale. In addition to these self-concept scores, reading and mathematics achievement scores and academic motivation scores were also collected as part of the study.

Results and Discussion

Factor analyses. Factor analyses of responses to the SDQ-I and the PCS both identified the factors that each instrument was designed to measure (see Table 25).

For SDQ-I responses the target coefficients (factor loadings of items designed to infer each factor) were consistently large (.33 to .92; median = .77), whereas nontarget loadings (all other factor loadings) were much smaller (-.11 to .24; median = .04). Similarly, for PCS responses, the target coefficients are consistently large (.32 to .75; median = .56), whereas nontarget loading are much smaller (-.14 to .34; median = .04). These results replicate previous factor analyses of responses to each of the self-concept instruments.

Table 25. Summary of Factor Analyses of Responses to the SDQ-I and the Perceived Competence Scale (PCS)

	Number of Coefficients	Highest Coefficients	Lowest Coefficients	Median Coefficients
SDQ-I				
Target Loadings	32	.92	.33	.77
Non-target Loadings	224	.24	-.11	.04
Communalities	32	.91	.54	.72
Factor Correlations	28	.46	.06	.19
PCS				
Target Loadings	28	.65	.32	.56
Non-target Loadings	84	.34	-.14	.04
Communalities	28	.57	.25	.41
Factor Correlations	6	.35	.07	.31

Note: Target loadings are the factor loadings of each variable on the one factor that it was designed to measure. All other factor loadings are non-target loadings. Factor correlations are the factor pattern correlations among the oblique factors identified in each analysis. Communalities are one minus the error variance (random error and specific variance) associated with each variable in the two analyses.

Multitrait-multimethod analyses. For purposes of this multitrait-multimethod analysis, only matching PCS (Physical, General, and Cognitive) and SDQ-I (Physical Abilities, Peer Relations, General-Self, and Total Academic) scores are considered (see Table 26). In applying the four criteria developed by Campbell and Fiske (1959; Marsh, in press), it was found that:

- 1) the four convergent validities (in bold italics) were all statistically significant and substantial (mean $r = .65$);
- 2) convergent validities (mean $r = .65$) were higher than other correlations in the same row and same column of the square (heterotrait-heteromethod) submatrix relating PCS and SDQ-I responses (mean $r = .30$) for all 24 comparisons, thus supporting this aspect of discriminant validity;
- 3) convergent validities were higher than other (heterotrait-heteromethod) correlations among PCS scales (mean $r = .41$) and among SDQ-I scores (mean $r = .45$) for 23 of 24 comparisons, thus supporting this aspect of discriminant validity; and
- 4) the pattern of correlations among PCS and SDQ-I scores was similar, suggesting that the pattern is independent of the instrument.

Correlations involving the remaining SDQ-I scores (General-School, Reading, Mathematics, Physical Appearance, and Parent Relations), though not formally considered as part of the multitrait-multimethod analysis, also supported the multitrait-multimethod findings: (a) the SDQ-I General-School, Reading, and Math scores were most substantially correlated with the PCS Cognitive score (.40 to .54), less correlated with the General scores from each instrument (.22 to .34), and even less correlated with the remaining scales (.01 to .33); (b) the SDQ-I Physical

Table 26. Correlations Between SDQ-I and Perceived Competence Scale (PCS), and Academic Motivation Measures in Reading (RAch) and Mathematics (MAch)

	PCS				SDQ-I ^a				SDQ-I ^b				SDI		Achievement	
	Physical	Social	General	Cognitive	Physical Abilities	Peer Relations	General-Self	Total Academic	General-School	Reading	Math	Physical Appearance	Parent Relations	Achievement Motivation	Reading Achievement	Math Achievement
PCS																
Phy	.82															
Soc	.46	.83														
Gen	.52	.50	.80													
Cog	.33	.15	.51	.79												
SDQ-I																
Phy	.67	.34	.36	.18	.87											
Peer	.41	.74	.45	.07	.46	.89										
Gen	.48	.38	.57	.35	.55	.66	.86									
T.Ach	.18	.07	.35	.60	.33	.26	.45	.93								
Sch	.16	.08	.34	.54	.33	.26	.43	.94	.89							
Read	.14	.01	.22	.46	.21	.15	.31	.72	.43	.94						
Math	.19	.12	.28	.40	.23	.27	.32	.55	.62	.16	.93					
Apr	.41	.29	.50	.22	.53	.48	.64	.29	.30	.14	.18	.91				
Prt	.14	.23	.37	.17	.25	.31	.40	.33	.33	.12	.19	.19	.89			
SDI																
AMot	.08	.01	.26	.35	.13	.11	.21	.48	.48	.28	.36	.13	.17	.75		
Achievement																
RAch	-.08	-.06	.03	.35	-.11	-.05	.01	.28	.14	.43	.03	-.16	-.11	.05	.93	
MAch	.05	-.04	.19	.40	-.03	-.04	.08	.26	.26	.15	.31	-.05	-.01	.05	.47	.88

Note: Decimals are omitted. Correlations greater than .08 and .11 are statistically significant at $p < .05$ and $p < .01$, respectively. Internal reliability estimates (coefficient alphas) are in boldface. Correlations (convergent validities) between scores from different instruments representing the same content domain are in bold italics.

^aSDQ-I scores posited to match the four PCS scores are the basis of the multitrait-multimethod analysis.

^bSDQ-I scores not included in the multitrait-multimethod analysis.

Appearance score was most highly correlated with the General-Self scale followed by the Physical Abilities and Social scales for both instruments; and (c) the SDQ-I Parent Relations scale was most highly correlated with the General scales for both instruments.

In summary, these results provide strong support for both the convergent and discriminant validity of responses to these two multidimensional self-concept instruments. These results also differ dramatically from those of the classic multitrait-multimethod studies conducted by Marx and Winne (1978; Winne, Marx, & Taylor, 1977) that were based on other instruments. The different results, as anticipated by Harter (1983), are apparently due to using two self-concept instruments in which the items are more carefully constructed with respect to their domain specificity.

Additional tests of construct validity. Correlations between the academic motivation and self-concept scores indicate that students with a more intrinsic orientation have higher self-concepts. For both PCS and SDQ-I, academic motivation scores were most substantially correlated with academic self-concept measures (.35 and .48), less correlated with general self-concept (.26 and .21), and relatively uncorrelated with physical and social self-concepts (.01 to .13). The motivation score was more highly correlated with the SDQ-I General-School (.48) and Total Academic scores (.48) than the PCS Academic scale (.35). These results provide clear support for the convergent and discriminant validity of the self-concept responses with respect to this academic motivation measure. Somewhat surprisingly, however, the motivation measure was not significantly correlated with reading or mathematics achievement scores. This indicates that motivation/self-concept relations are independent of the achievement/self-concept relations.

Reading and math achievement scores were most highly correlated with the PCS Academic score and less highly correlated with the other PCS scores. Reading achievement was most highly correlated with the SDQ-I Reading score, less correlated with the SDQ-I General-School and Total Academic scores, uncorrelated with the SDQ-I Math score, and uncorrelated or negatively correlated with the remaining SDQ-I scores. Math achievement was most highly correlated with the SDQ-I Math score, less correlated with the SDQ-I General-School and Total Academic scores, still less correlated with the SDQ-I Reading score, and not significantly correlated with the remaining SDQ-I scores. Results for both the PCS and SDQ-I support the convergent and discriminant validity of the domain specific measures of self-concept with respect to academic achievement, but the SDQ-I results further support the separation of the Reading and Math self-concepts as emphasized in the revision of the Shavelson model (Marsh, Byrne, & Shavelson, 1988; Marsh & Shavelson, 1985).

In summary, results of factor analyses, multitrait-multimethod analyses, and correlations with other criterion measures all provided support for the convergent and divergent validity of responses to both self-concept instruments. Recent research by Byrne (1988) provides further substantiation of these findings. Byrne collected responses to the PCS and the SDQ-I for small samples of gifted children (IQs of 129 and above) in grades 5 ($N = 44$) and 8 ($N = 46$) attending public schools in Ottawa, Canada. Though not the focus of that study, Byrne reported convergent validities ranging from .54 to .86 (mean $r = .73$) for grade 5, and from .66 to .88 (mean $r = .80$) for grade 8. Based on the present findings and those summarized by Byrne, use of either the PCS or the SDQ-I self-concept instruments appears to be warranted.

The distinctive features of the SDQ-I and PCS instruments are the additional scales and added length of the SDQ-I, and the alternative response format of the PCS. (On the PCS each "item" consists of two logically opposed statements so that children

must first make a "forced choice" selection of which statement is most appropriate and then rate the extent to which the selected statement applies to them.) Marsh (1986d; Marsh, Byrne, & Shavelson, in press) provided convincing evidence for the separation of the Reading and Math self-concepts and argued that academic self-concept cannot be adequately understood if only a general academic scale is considered. Clear support for this separation of Math and Verbal self-concepts was also found in this study. Thus, researchers interested in separate estimates of reading or math self-concepts should use the SDQ-I. Similarly, if researchers are interested in measures of physical appearance or parent relations self-concepts, then the SDQ-I is recommended. Whereas the SDQ-I has three times as many items as the PCS, the fact that each PCS "item" actually consists of two statements largely offsets this difference. In the present investigation there was little difference in the time required to complete the two instruments. Harter (1982) suggested that the PCS response format reduces social desirability responding, which may be an important advantage of the PCS. However, the author of the SDQ knows of no empirical support for this suggestion that is based on comparisons of responses to the same items using a standard and the alternative response format. Furthermore, particularly for younger children or less intelligent children, the format may be confusing (also see Chapter 13 on the use of negatively worded items with young children) and the PCS factor structure is not as well defined for these groups. Whereas the PCS may be inappropriate for these groups (e.g., Harter, 1982; Silon & Harter, 1985), it is not clear whether this problem is inherent in the measurement of self-concept or a function of the idiosyncratic PCS response scale. Hence, whereas Harter claims that the alternative response format is a positive feature of the PCS, further evaluation is needed, particularly for younger and less intelligent children.

Summary and Implications

One widely accepted approach to establishing construct validity is to correlate scores from different instruments designed to measure the same construct. Particularly for multidimensional constructs, this approach is facilitated by the application of multitrait-multimethod analysis. Despite the popularity of this approach, it has not been applied frequently in self-concept research, and until recently the few applications that were made have not been supportive of the construct validity of self-concept responses. The reason, apparently, is that prior to the 1970s and even the 1980s, self-concept instruments have not been designed specifically to measure a priori components of self-concept and did not have well-defined factor structures. If there is ambiguity in the factor structure of individual instruments, then it is not surprising that there is ambiguity in relations between responses from different instruments. In contrast to most self-concept instruments, the SDQ instruments were designed specifically to measure a priori factors that have been identified in many factor analytic studies. For this reason, a logical pattern of relations between SDQ responses and those from other multidimensional instruments is more likely than in previous research such as the Marx and Winne (1978) study. Support presented here and in the other SDQ manuals provides clear support for this expectation.

Chapter 15.

Summary of Theoretical and Empirical Research

Interest in self-concept has resulted from the recognition that it is 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 the interest in how self-concept is related to other variables, and from an 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. There are 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, Hubner, & Stanton, 1976; Wells & Marwell, 1976; Wylie, 1974, 1979) typically emphasize the lack of a clear 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. The disappointing lack of rigor can perhaps be explained by the failure to identify self-concept research with any particular discipline.

Theoretical Basis

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 model, as well as their structure, were heuristic and plausible, they were not supported by empirical research. In particular, Shavelson et al. were unable to identify any existing instruments which measured the facets posited in their model, and not even the multidimensionality of self-concept was 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 instrumentation employed in self-concept research. It was clear that development of a suitable instrument was necessary for empirically testing the model and for examining further theoretical issues.

This need provided the initial impetus for the development and refinement of the SDQ instruments. 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 used as the starting point for instrument construction, and empirical results were used to support, refute, or revise the instrument and the theory upon 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-I is based upon strong empirical and theoretical bases.

Construct Validation

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: whether self-concept is a unidimensional or multidimensional construct, and if multidimensional, what dimensions characterize it. These studies typically employ factor analysis and multitrait-multimethod analysis. Between-network studies attempt to demonstrate a theoretically consistent, or at least logical, pattern of relationships between measures of self-concept and other constructs.

SDQ-I research includes both within-network and between-network studies. Early SDQ-I research, primarily within-network studies, focused on internal characteristics of self-concept, particularly its facets and their organization. More recent SDQ-I research, between-network studies, examined 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-I and the Shavelson model upon which it is based. The research also clarifies many theoretical issues in self-concept research. In this sense SDQ-I research represents an interplay between theory and empirical research and supports the construct validity approach which has guided SDQ-I research.

Perhaps the strongest contribution of the SDQ-I research to the measurement of self-concept is the description of a well-developed instrument, based on a strong empirical foundation and a good theoretical model. Reviewers in this field typically argue that the most important weakness in self-concept research is the poor quality of measurement instruments, and it is anticipated that the SDQ-I will help to remedy this situation.

References

- Alwin, D. F., & Jackson, D. J. (1981). Applications of simultaneous factor analysis to issues of factorial invariance. In D. J. Jackson & E. F. Borgotta (Eds.), *Factor analysis and measurement in sociological research: A multidimensional perspective*. Beverly Hills, CA: Sage, pp. 249-280.
- Andrews, R. J., Elkin, J., & Cochrane, R. G. (1972). *Manual for the Moreton Mathematics Tests — Levels II & III metric revision*. Brisbane, Australia: Teaching and Testing Resources.
- Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Princeton, NJ: Educational Testing Service.
- Bachman, J. G. (1970). *Youth in transition (Vol. 2): The impact of family background and intelligence on tenth-grade boys*. Ann Arbor: Institute for Social Research, The University of Michigan.
- 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.
- Bachman, J. G., & O'Malley, P. M. (1984). Black-white differences in self-esteem: Are they due to differences in response style? *American Journal of Sociology*, 90, 624-639.
- Bachman, J. G., & O'Malley, P. M. (1986). Self-concepts, self-esteem, and educational experiences: The frog pond revisited (again). *Journal of Personality and Social Psychology*, 50, 33-46.
- Bachman, J. G., O'Malley, P. M., & Johnston, J. (1978). *Youth in transition (Vol. 4): Adolescence to adulthood — change and stability in the lives of young men*. Ann Arbor: Institute for Social Research, The University of Michigan.
- Bar-Tal, D. (1978). Attributional analysis of achievement-related behavior. *Review of Educational Research*, 48, 259-271.
- Benson, J., & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. *Journal of Educational Measurement*, 22, 231-240.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Boersma, F. J., & Chapman, J. W. (1979). *Student's Perception of Ability Scale Manual*. Edmonton, Canada: University of Alberta.
- Bradley, G. W. (1978). Self-serving biases in the attribution process: A reexamination of the fact or fiction question. *Journal of Personality and Social Psychology*, 36, 56-71.
- Brookover, W. B., & Erikson, E. L. (1975). *Sociology of education*. Homewood, IL: Dorsey Press.

Brookover, W. B., & Passalacqua, J. (1981). Comparison of aggregate self-concepts for populations with different reference groups. In M. Lynch, A. Norem-Hebeisen, & W. Gergen (Eds.), *Self-concept: Advances in theory and research*. Cambridge, MA: Ballinger.

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. (1988). The Self-Description Questionnaire-III: Testing for equivalent factorial validity across ability. *Educational and Psychological Measurement*, 48, 397-406.

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

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

Combs, A., Soper, D., & Courson, C. (1963). The measurement of self-concept and self-report. *Educational and Psychological Measurement*, 23, 493-500.

Connell, J. P., & Ryan, R. M. (1984). A developmental theory of motivation in the classroom. *Teacher Education Quarterly*, 11, 67-77.

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

Covington, M. V. (1984). The motive of self-worth. In R. E. Ames & C. Ames (Eds.), *Research on motivation in education* (Vol. 1, pp. 78-114). Orlando, FL: Academic Press.

Covington, M. V., & Omelich, C. L. (1979). Are causal attributions causal? A path analysis of the cognitive model of achievement motivation. *Journal of Personality and Social Psychology*, 37, 1487-1504.

Crandall, R. (1973). The measurement of self-esteem and related constructs. In J. Robinson & P. Shaver (Eds.), *Measures of social psychological attitudes* (pp. 45-158). Ann Arbor: Institute for Social Research, The University of Michigan.

Critelli, J. W., & Neumann, K. F. (1984). The placebo: Conceptual analysis of a construct in transition. *American Psychologist*, 39, 32-39.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.

Daly, J. A., Bell, R. A., & Korinek, J. (1987). Interrelationships among attitudes toward academic subjects. *Contemporary Educational Psychology*, 12, 147-155.

Davis, J. A. (1966). The campus as a frog pond: An application of theory of relative deprivation to career decisions for college men. *American Journal of Sociology*, 72, 17-31.

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).

Dyer, C. O. (1964). Construct validity of self-concept by a multitrait-multimethod analysis (Doctoral dissertation, University of Michigan, 1963). *Dissertation Abstracts International*, 25, 8154.

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.

Fennema, E. (1974). Mathematics learning and the sexes: A review. *Journal for Research in Mathematics Education*, 5, 126-139.

Fernandes, M. W., Michaels, W. B., & Smith, R. (1978). The factorial validity of three forms of the *Dimensions of Self-Concept* measure. *Educational and Psychological Measurement*, 38, 537-545.

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.

Fleming, J. S., & Watts, W. A. (1980). The dimensionality of self-esteem: Results for a college sample. *Journal of Personality and Social Psychology*, 39, 921-929.

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

Gottfried, A. E. (1982). Relations between academic intrinsic motivation and anxiety in children and young adolescents. *Journal of School Psychology*, 20, 205-215.

Gottfried, A. E. (1985). Academic intrinsic motivation in elementary and junior high school students. *Journal of Educational Psychology*, 77, 631-645.

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.

Harvey, J. H., & Weary, G. (1984). Current issues in attribution theory and research. *Annual Review of Psychology*, 35, 427-459.

Hays, W. L. (1981). *Statistics* (3rd Ed.). New York: Holt, Rinehart, & Winston, pp. 210-211.

- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Hull, C. H., & Nie, N. H. (1981). *SPSS Update 7-9*. New York: McGraw-Hill.
- Jackson, S. A., & Marsh, H. W. (1986). Athletic or antisocial: The female sport experience. *Journal of Sport Psychology*, 8, 196-211.
- Jerusalem, M. (1984). Reference group, learning environment and self-evaluations: A dynamic multi-level analysis with latent variables. In R. Schwarzer (Ed.), *The self in anxiety, stress and depression*, (pp.61-73). North-Holland: Elsevier Science Publishers.
- Joreskog, K. G., & Sorbom, D. (1981). *LISREL V: Analysis of linear structural relations by the method of maximum likelihood*. Chicago: International Educational Services.
- Kinch, J. W. (1963). A formalized theory of self-concept. *American Journal of Sociology*, 68, 481-486.
- Kulik, C. L. (1985). *Effects of inter-class ability grouping on achievement and self-esteem*. Paper presented at the 1985 Annual Meeting of the American Psychological Association, Los Angeles.
- Kulik, C. L., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal*, 21, 799-806.
- Marsh, H. W. (1984a). Relationships among dimensions of self-attribution, dimensions of self-concept, and academic achievements. *Journal of Educational Psychology*, 76, 1291-1308.
- Marsh, H. W. (1984b). Self-concept, social comparison and ability grouping: A reply to Kulik and Kulik. *American Educational Research Journal*, 21, 799-806.
- Marsh, H. W. (1984c). Self-concept: The application of a frame of reference model to explain paradoxical results. *Australian Journal of Psychology*, 28, 165-181.
- 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 phenomena. *Developmental Psychology*, 22, 37-49.
- Marsh, H. W. (1986b). Global self-esteem: Its relation to 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 factorial invariance of responses by males and females to a multidimensional self-concept instrument: Substantive and methodological issues. *Multivariate Behavioral Research*, 22, 457-480.
- Marsh, H. W. (1987c). The hierarchical structure of self-concept and the application of hierarchical confirmatory factor analysis. *Journal of Educational Measurement*, 24, 17-19.
- Marsh, H. W. (1988a). *The frame of reference effects of school-average ability: The High School and Beyond study*. Manuscript submitted for publication.
- Marsh, H. W. (1988b). *Multidimensional self-concepts: Do positively and negatively worded items measure substantively different components of self?* Manuscript submitted for publication.
- Marsh, H. W. (1988c). *Sex differences in the development of verbal and math constructs: The High School and Beyond study*. Manuscript submitted for publication.
- Marsh, H. W. (1988d). *Age and sex effects in multiple dimensions of self-concept: Preadolescence to early-adulthood*. Manuscript submitted for publication.
- Marsh, H. W. (in press). Multitrait-multimethod analyses. In J. P. Keeves (Ed.), *Educational research methodology, measurement and evaluation: An international handbook*. Oxford: Pergamon Press.
- 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., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103, 391-411.
- 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 & multitrait-multimethod analysis. *Journal of Personality and Social Psychology*, 49, 1360-1377.
- Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (in press). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*.
- Marsh, H. W., Cairns, L., Relich, J. D., Barnes, J., & Debus, R. (1984). The relationship between dimensions of self-attribution and dimensions of self-concept. *Journal of Educational Psychology*, 76, 3-32.
- Marsh, H. W., & Gouvenet (in press). Multidimensional self-concepts and perceptions of control: Construct validation of responses by children. *Journal of Educational Psychology*.

- Marsh, H. W., & Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. *Journal of Educational Measurement, 20*, 231-248.
- 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., McDonald, R. P., & Balla, J. R. (1987). *Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size and of over-fitting and under-fitting models*. Manuscript submitted for publication.
- 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 (SDQ-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 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., Parker, J. W., & Smith, I. D. (1983). Preadolescent self-concept: Its relation to self-concept as inferred by teachers and to academic ability. *British Journal of Educational Psychology, 53*, 60-78.
- Marsh, H. W., & Peart, N. (in press). Competitive and cooperative physical fitness training programs for girls: Effects on physical fitness and on multidimensional self-concepts. *Journal of Sport Psychology*.
- 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. (1988). Self-other agreement and self-other differences on multidimensional self-concept ratings. Manuscript submitted for publication.
- Marsh, H. W., & Richards, G. (in press). The Outward Bound Bridging Course for low achieving high-school males: Effect on academic achievement and multidimensional self-concepts. *Australian Journal of Psychology*.
- Marsh, H. W., Richards, G., & Barnes, J. (1986a). Multidimensional self-concepts: A long-term follow-up of the effect of participation in an Outward Bound program. *Personality and Social Psychology Bulletin, 12*, 475-492.
- Marsh, H. W., Richards, G., & Barnes, J. (1986b). Multidimensional self-concepts: The effect of participation in an Outward Bound program. *Journal of Personality and Social Psychology, 45*, 173-187.
- Marsh, H. W., & Shavelson, R. J. (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. (1987). A cross-national study of the structure and level of multidimensional self-concepts: An application of confirmatory factor analysis. *Australian Journal of Psychology, 39*, 61-77.
- Marsh, H. W., Smith, I. D., & Barnes, J. (1983). Multitrait-multimethod analyses of the Self-Description Questionnaire: Student-teacher agreement on multidimensional ratings of student self-concept. *American Educational Research Journal, 20*, 333-357.
- Marsh, H. W., Smith, I. D., & Barnes, J. (1984). Multidimensional self-concepts: Relationships with inferred self-concepts and academic achievement. *Australian Journal of Psychology, 36*, 367-386.
- Marsh, H. W., Smith, I. D., & Barnes, J. (1985). Multidimensional self-concepts: Relationships with sex and academic achievement. *Journal of Educational Psychology, 77*, 581-596.
- Marsh, H. W., Smith, I. D., Barnes, J., & Butler, S. (1983). Self-concept: Reliability, dimensionality, validity, and the measurement of change. *Journal of Educational Psychology, 75*, 772-790.
- Marsh, H. W., Smith, I. D., Myers, M. R., & Owens, L. (in press). The transition from single-sex to coeducational high schools: Effects on multiple dimensions of self-concept and on academic achievement. *American Educational Research Journal*.
- Maruyama, G., Rubin, R. A., & Kingsbury, G.G. (1981). Self-esteem and education: Independent constructs or a common cause. *Journal of Personality and Social Psychology, 40*, 962-975.
- Marx, R. W., & Winne, P. H. (1978). Construct interpretations of three self-concept inventories. *American Educational Research Journal, 15*, 99-108.
- McClelland, D. C. (1965). Toward a theory of motive acquisition. *American Psychologist, 20*, 321-333.
- McLeod, J. (1972). *GAPADOL Reading Comprehension Manual*. Richmond, Australia: Heinemann Educational.
- 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.
- Michaels, W. B., Smith, R. A., & Michaels, J. J. (1975). The factorial validity of the Piers-Harris Children's Self-concept Scale for each of three samples of elementary, junior high and senior high students in a large metropolitan school district. *Educational and Psychological Measurement, 35*, 405-414.
- Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin, 82*, 213-225.
- Miyamoto, S. F., & Dornbusch, S. (1956). A test of interactionist hypotheses of self-conception. *American Journal of Sociology, 15*, 399-403.
- Newman, R. S. (1980). Alleviating learned helplessness in a wilderness setting: An application of attribution theory to Outward Bound. In L. J. Fyans (Ed.), *Achievement motivation: Recent trends in theory and research* (pp. 312-345). New York: Plenum Press.

- Nicholls, J. (1979). Development of perception of attainment and causal attributions for success and failure in reading. *Journal of Educational Psychology, 71*, 94-99.
- Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. (1975). *Statistical Package for the Social Sciences*. New York: McGraw-Hill.
- Piers, E. V., & Harris, D. A. (1964). Age and other correlates of self-concept in children. *Journal of Educational Psychology, 55*, 91-95.
- Relich, J. D. (1983). *Attribution, affective variables and arithmetic achievement*. Unpublished dissertation. The University of Sydney, Australia.
- 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.
- Riess, M., Rosenfield, P., Melburg, V., & Tedeschi, J. T. (1981). Self-serving attributions: Biased private perceptions and distorted public dispositions. *Journal of Personality and Social Psychology, 41*, 224-231.
- Rogers, C. M., Smith, M. D., & Coleman, J. M. (1978). Social comparison in the classroom: The relationship between academic achievement and self-concept. *Journal of Educational Psychology, 70*, 50-57.
- Rosenberg, M. (1965). *Society and the adolescent child*. Princeton, NJ: Princeton University Press.
- Rosenberg, M. (1979). *Conceiving the self*. New York: Basic Books.
- 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., & Loebel, J. H. (1980). Developmental analysis of the role of social comparison in self-evaluations. *Developmental Psychology, 16*, 105-115.
- Ryan, R. M., & Connell, J. P. (1988). Perceived locus of causality and internalization: A reason simplex model. Manuscript submitted for publication.
- Ryan, R. M., Connell, J. P., & Grolnick, W. S. (in press). When achievement is not intrinsically motivated: A theory and assessment of self-regulation in school. In A. K. Boggiano & T. S. Pittman (Eds.), *Achievement and motivation: A social-developmental perspective*.
- Scheirer, M. A., & Kraut, R. E. (1979). Increasing educational achievement via self-concept change. *Review of Educational Research, 49*, 131-150.
- Schwarzer, R., Jerusalem, J., & Lange, B. (1983). *The change of self-concept with respect to reference groups in school*. Paper presented at the 1983 Annual Meeting of the American Educational Research Association, Montreal.
- Sears, P. S. (1963). *The effect of classroom conditions on the strength of achievement motive and work output of elementary school children*. USOE Cooperative Research Project, Project No. OE-873, Stanford, CA: Stanford University.

- 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*. Hillsdale, NJ: Lawrence Erlbaum.
- Shepard, L. A. (1979). Self-acceptance: The evaluative components of the self-concept construct. *American Educational Research Journal, 16*, 139-160.
- Sherman, J. (1980). Mathematics, spatial visualization, and related factors: Changes in boys and girls, grades 8-11. *Journal of Educational Psychology, 72*, 476-482.
- Shibutani, T. (1961). *Society and personality*. Englewood Cliffs, NJ: Prentice-Hall.
- Shore, A. (1977). *Outward Bound: A reference volume*. Greenwich, CT: Outward Bound, Inc.
- Shrauger, J. S., & Schoeneman, T. J. (1979). Symbolic interactionist view of self-concept: Through the glass darkly. *Psychological Bulletin, 86*, 549-573.
- Silon, E. L., & Harter, S. (1985). Assessment of perceived competence, motivational orientation, and anxiety in segregated and mainstreamed educable mentally retarded children. *Journal of Educational Psychology, 77*, 217-230.
- Smith, I. D., & Marsh, H. W. (1985). A cross-cultural study of self-concept in primary school children. *Collected Papers of the 1985 Annual Conference of the Australian Association for Research in Education*, pp. 350-353.
- Smith, M. L., Gabriel, R., Schott, J., & Padia, W. L. (1975). *Evaluation of the effects of Outward Bound*. Boulder: Bureau of Educational Field Services, University of Colorado.
- Snyder, M. L., Stephan, W. G., & Rosenfield, D. (1978). Attributional Egotism. In J. H. Harvey, W. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research. Vol. 2*, (pp. 91-117). Hillsdale, NJ: Lawrence Erlbaum.
- Soares, A. T., & Soares, L. M. (1969). Self-perceptions of culturally disadvantaged children. *American Educational Research Journal, 6*, 31-45.
- Soares, L. M., & Soares, A. T. (1977). *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). *Convergence and discrimination in academic self-concepts*. Paper presented at the 20th Congress of the International Association of Applied Psychology, Edinburgh, Scotland.
- 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.

Stipek, D. J., & Weisz, J. R. (1981). Perceived personal control and academic achievement. *Review of Educational Research*, 51, 101-137.

Strang, L., Smith, M. D., & Rogers, C. M. (1978). Social comparison, multiple reference groups and the self-concepts of academically handicapped children before and after mainstreaming. *Journal of Educational Psychology*, 70, 487-497.

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

Trowbridge, N. (1970). Effects of socio-economic class on self-concept of children. *Psychology in the Schools*, 7, 304-306.

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

Weiner, B. (1972). *Theories of motivation: From mechanism to cognition*. Chicago: Rand McNally.

Weiner, B. (1974). *Achievement motivation and attribution theory*. Morristown, N. J.: General Learning Press.

Weiner, B. (1980). *Human motivation*. New York: Holt, Rinehart, & Winston.

Weiner, B. (1983). Some methodological pitfalls in attributional research. *Journal of Educational Psychology*, 75, 530-543.

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

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

Winne, P. H., Marx, R. W., & Taylor, T. D. (1977). A multitrait-multimethod study of three self-concept inventories. *Child Development*, 48, 893-901.

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.

Zuckerman, M. (1979). Attribution of success and failure revisited, or: The motivational bias is alive and well in attribution theory. *Journal of Personality*, 47, 245-287.

Appendices

Appendix A-1. Percentiles and T Scores for the Total Normative Sample (N = 3,562)

Raw Score	Scale										Total Academic %	Total Nonacademic %	General-School %	General-Self %	Total Self %	Raw Score
	Physical Abilities %	Physical Appearance %	Peer Relations %	Parent Relations %	Reading %	Mathematics %	General-School %	General-Self %	Total Academic %	Total Nonacademic %						
8	11	1	27	15	19	1	26	21	5	2	15	1	8			
9	12	2	28	16	20	2	28	22	7	4	17	3	9			
10	14	3	29	18	22	3	29	24	9	6	18	5	10			
11	16	4	31	19	23	4	30	25	11	8	20	8	11			
12	17	5	32	21	24	5	31	27	13	10	22	10	12			
13	19	6	33	22	26	6	32	28	14	12	23	12	13			
14	20	8	34	24	27	7	33	29	16	14	25	14	14			
15	22	9	35	25	28	8	34	31	18	16	26	16	15			
16	23	11	36	27	29	10	36	32	20	18	28	18	16			
17	25	13	38	28	31	12	37	34	22	20	30	20	17			
18	27	16	39	30	32	14	38	35	23	22	31	23	18			
19	28	18	40	32	34	16	39	36	25	24	33	25	19			
20	30	20	41	33	35	18	40	38	27	26	35	27	20			
21	31	22	42	35	36	20	41	39	29	28	36	29	21			
22	33	25	44	36	38	22	42	41	30	30	38	30	22			
23	35	28	45	38	39	24	43	42	32	32	39	32	23			
24	36	32	46	40	40	26	44	44	34	34	41	34	24			
25	38	37	47	41	42	28	46	46	36	36	43	36	25			
26	39	41	48	42	43	30	47	48	38	38	44	38	26			
27	41	46	49	44	44	32	48	49	40	40	46	40	27			
28	43	49	51	45	46	34	49	51	42	42	48	42	28			
29	44	53	52	47	47	36	50	52	44	44	50	44	29			
30	46	57	53	48	48	38	51	54	46	46	52	46	30			
31	47	61	54	49	49	40	52	55	48	48	54	48	31			
32	49	65	55	50	50	42	53	57	50	50	56	50	32			
33	51	69	56	53	51	44	54	58	52	52	58	52	33			
34	52	73	58	55	52	46	55	59	54	54	60	54	34			
35	54	77	59	56	53	48	56	60	56	56	62	56	35			
36	55	81	60	58	54	50	57	61	58	58	64	58	36			
37	57	84	61	59	55	52	58	62	60	60	66	60	37			
38	59	88	62	61	57	54	59	63	62	62	68	62	38			
39	60	92	64	62	58	56	60	64	64	64	70	64	39			
40	62	97	65	64	60	58	62	66	66	66	74	66	40			
Mean	32.65	27.49	31.01	35.45	31.28	28.78	28.55	32.99	31.77	31.77	29.53	30.89	Mean			
Median	33.93	28.13	31.96	36.93	32.73	29.94	29.22	33.98	32.32	32.32	29.93	31.18	Median			
SD	6.28	8.48	6.51	5.13	7.48	8.83	7.05	5.60	4.96	4.96	6.17	4.69	SD			
Skewness	-.91	-.41	-.77	-1.75	-.82	-.51	-.48	-1.23	-.55	-.55	-.47	-.38	Skewness			
Reliability	.83	.90	.85	.80	.89	.92	.86	.81	.91	.91	.92	.94	Reliability			
r with year	-.11	-.20	-.10	-.07	-.14	-.11	-.17	-.17	-.18	-.18	-.17	-.21	r with year			
r with sex	-.29	-.11	-.09	-.02	.16	-.12	.04	-.12	-.17	-.17	.02	-.08	r with sex			

Note: r with year not shown for the General-Self scale; only the newest version of the SDQ-I contains this scale. *N = 1,118

Appendix A-4. Percentiles and T Scores: Males, Grades 5-6 (N = 1,583)

Raw Score	Scale												Total Academic %	Total Nonacademic %	General-Self %	Total %	Raw Score		
	Physical Abilities		Physical Appearance		Peer Relations		Parent Relations		Reading		Mathematics							General-School	
	T	%	T	%	T	%	T	%	T	%	T	%						T	%
8	3	1	25	12	1	26	23	3	1	26	23	3	17	18	3	8			
9	5	2	26	14	1	27	24	4	1	27	24	4	1	18	3	9			
10	6	2	28	16	1	28	25	6	1	28	25	6	2	20	5	10			
11	8	3	29	17	1	29	26	8	1	29	26	8	4	21	7	11			
12	10	3	30	19	1	30	27	10	1	30	27	10	7	23	9	12			
13	12	4	31	20	1	31	28	12	3	31	28	12	9	25	11	13			
14	14	6	33	22	1	33	30	14	4	33	30	14	11	26	13	14			
15	16	7	34	24	1	34	31	16	5	34	31	16	13	28	16	15			
16	17	9	35	25	1	35	32	17	6	35	32	17	15	29	18	16			
17	19	11	36	27	1	36	33	19	7	36	33	19	17	31	20	17			
18	21	13	37	28	1	37	34	21	8	37	34	21	19	32	22	18			
19	23	15	39	30	1	39	35	23	9	39	35	23	21	34	24	19			
20	25	17	40	32	1	40	36	25	10	40	36	25	24	36	26	20			
21	26	19	41	33	2	41	37	26	11	41	37	26	26	37	28	21			
22	28	21	42	35	2	42	38	28	12	42	38	28	28	39	31	22			
23	30	23	44	37	3	44	39	30	13	44	39	30	30	40	33	23			
24	32	25	45	38	4	45	40	32	14	45	40	32	32	42	35	24			
25	34	27	46	40	4	46	41	34	15	46	41	34	34	44	37	25			
26	35	29	47	41	5	47	42	35	16	47	42	35	36	45	39	26			
27	37	31	48	43	5	48	43	37	17	48	43	37	39	47	41	27			
28	39	33	49	44	6	49	44	39	18	49	44	39	41	48	44	28			
29	41	35	50	45	6	50	45	41	19	50	45	41	43	50	46	29			
30	43	37	51	46	7	51	46	43	20	51	46	43	45	53	48	30			
31	45	39	52	47	7	52	47	45	21	52	47	45	47	55	50	31			
32	47	41	53	48	8	53	48	47	22	53	48	47	49	57	52	32			
33	49	43	54	49	8	54	49	49	23	54	49	49	51	60	54	33			
34	51	45	55	50	9	55	50	51	24	55	50	51	53	62	56	34			
35	53	47	56	51	9	56	51	53	25	56	51	53	55	64	58	35			
36	55	49	57	52	10	57	52	55	26	57	52	55	57	66	60	36			
37	57	51	58	53	10	58	53	57	27	58	53	57	60	68	62	37			
38	59	53	59	54	11	59	54	59	28	59	54	59	62	70	64	38			
39	61	55	60	55	11	60	55	61	29	60	55	61	64	72	66	39			
40	63	57	61	56	12	61	56	63	30	61	56	63	66	74	67	40			
Mean	34.11	28.16	31.42	35.48	29.96	29.44	27.85	33.45	27.85	29.44	27.85	33.45	32.39	29.07	30.96	Mean			
Median	35.65	28.87	32.52	36.89	31.26	31.04	28.35	34.24	28.35	31.04	28.35	34.24	32.95	29.43	31.32	Median			
SD	5.54	8.12	6.24	5.04	7.83	8.92	7.28	5.37	7.28	8.92	7.28	5.37	4.70	6.32	4.63	SD			
Skewness	-1.14	-0.48	-0.85	-1.73	-0.66	-0.61	-0.42	-1.46	-0.42	-0.61	-0.42	-1.46	-0.65	-0.47	-0.44	Skewness			
Reliability	.80	.89	.85	.79	.89	.92	.87	.80	.87	.92	.87	.80	.91	.92	.93	Reliability			
r with year	-.08	-.17	-.09	-.11	-.09	-.08	-.10	-.07	-.10	-.08	-.10	-.07	-.17	-.11	-.17	r with year			

*n = 559

Appendix A-5. Percentiles and T Scores: Females, Grades 5-6 (N = 1,185)

Raw Score	Scale												Total Academic %	Total Nonacademic %	General-Self %	Total %	Raw Score		
	Physical Abilities		Physical Appearance		Peer Relations		Parent Relations		Reading		Mathematics							General-School	
	T	%	T	%	T	%	T	%	T	%	T	%						T	%
8	16	1	30	16	1	27	18	12	1	27	18	12	4	12	8	8			
9	18	3	31	17	1	28	19	13	1	28	19	13	6	14	9	9			
10	19	5	32	19	1	29	20	14	1	29	20	14	8	16	10	10			
11	21	6	33	20	1	30	21	15	1	30	21	15	10	17	11	11			
12	22	8	34	22	1	31	22	16	1	31	22	16	12	19	12	12			
13	24	10	35	23	2	32	23	17	1	32	23	17	14	21	13	13			
14	25	12	37	25	2	33	24	18	1	33	24	18	16	23	14	14			
15	27	14	38	27	1	34	25	19	1	34	25	19	18	24	15	15			
16	28	16	39	28	1	35	26	20	1	35	26	20	20	26	16	16			
17	30	18	40	30	1	36	27	21	1	36	27	21	22	28	18	17			
18	31	20	41	31	2	37	28	22	1	37	28	22	24	30	19	18			
19	33	22	42	33	2	38	29	23	1	38	29	23	26	32	20	19			
20	34	25	44	34	2	39	30	24	1	39	30	24	28	34	21	20			
21	36	28	45	36	3	40	31	25	1	40	31	25	30	36	22	21			
22	37	32	46	37	3	41	32	26	1	41	32	26	33	38	23	22			
23	39	36	47	39	4	42	33	27	1	42	33	27	35	40	24	23			
24	40	41	48	40	5	43	34	28	1	43	34	28	37	42	25	24			
25	42	46	49	42	6	44	35	29	1	44	35	29	39	44	26	25			
26	43	50	51	43	7	45	36	30	1	45	36	30	41	46	27	26			
27	45	55	52	45	8	46	37	31	1	46	37	31	43	48	28	27			
28	46	59	53	46	9	47	38	32	1	47	38	32	45	50	29	28			
29	48	62	54	48	10	48	39	33	1	48	39	33	47	52	30	29			
30	49	66	55	49	11	49	40	34	1	49	40	34	49	54	31	30			
31	51	70	56	51	12	50	41	35	1	50	41	35	51	56	32	31			
32	53	74	57	53	13	51	42	36	1	51	42	36	53	58	33	32			
33	54	78	59	54	14	52	43	37	1	52	43	37	55	60	34	33			
34	56	82	60	56	15	53	44	38	1	53	44	38	57	62	35	34			
35	57	85	61	57	16	54	45	39	1	54	45	39	59	64	36	35			
36	59	87	62	59	17	55	46	40	1	55	46	40	61	66	37	36			
37	60	90	63	60	18	56	47	41	1	56	47	41	63	68	38	37			
38	62	92	64	62	19	57	48	42	1	57	48	42	65	70	39	38			
39	63	94	66	64	20	58	49	43	1	58	49	43	67	72	40	39			
40	65	96	67	66	21	59	50	44	1	59	50	44	69	74	41	40			
Mean	30.34	25.56	30.18	35.30	32.28	27.33	28.54	31.94	28.54	32.28	27.33	31.94	30.47	29.38	30.16	Mean			
Median	31.23	25.90	30.67	36.95	33.81	27.88	29.02	33.50	29.02	33.81	27.88	33.50	30.77	29.69	30.41	Median			
SD	6.61	8.59	6.47	5.31	6.84	8.48	6.43	6.27	6.43	6.84	8.48	6.27	4.85	5.62	4.36	SD			
Skewness	-.60	-.24	-.62	-1.69	-.84	-.35	-.51	-.95	-.51	-.84	-.35	-.95	-.36	-.44	-.37	Skewness			
Reliability	.82	.91	.86	.83	.88	.90	.86	.84	.86	.88	.90	.84	.91	.92	.93	Reliability			
r with year	-.12	-.24	-.12	-.13	-.10	-.05	-.13	-.07	-.13	-.10	-.05	-.07	-.15	-.23	-.12	r with year			

*n = 559

Appendix B. Percentiles and T-Score Equivalents for the Experimental Control Scores

Raw Score	Control Score 1		Control Score 2		Control Score 3		Control Score 4		Control Score 5		Control Score 6		
	%	T	Raw Score	%	Raw Score	%	T	Raw Score	%	Raw Score	%	T	
0	98	71	0	22	≤-10	23	0	99	73	≤-19	23	≤0.3	23
1	97	69	1	26	-9	26	1	98	70	-18	26	0.4	26
2	95	67	2	28	-8	26	2	96	68	-17	26	0.5	26
3	94	66	3	29	-7	26	3	94	67	-16	27	0.6	26
4	92	64	4	30	-6	26	4	91	64	-15	27		26
5	89	62	5	31	-5	26	5	87	61	-14	28	0.7	26
6	84	60	6	32	-4	28	6	80	58	-13	29	0.8	26
7	78	58	7	33	-3	30	7	72	56	-12	30	0.9	26
8	73	56	8	34	-2	32	8	63	53	-11	32	1.0	27
9	67	54	9	35	-1	33	9	53	51	-10	34	1.2	28
10	58	51	10	36	0	34	10	43	48	-9	36	1.4	29
11	50	49	11	37	0	36	11	34	46	-8	37	1.6	30
12	44	49	12	37	1	37	12	27	44	-7	39	1.8	32
13	40	47	13	38	2	38	13	21	42	-7	41	2.0	33
14	34	46	14	39	3	39	14	16	40	-6	43	2.2	34
15	27	44	15	41	4	40	15	12	38	-5	45	2.4	35
16	21	42	16	42	5	42	16	8	37	-4	47	2.6	36
17	18	41	17	42	6	43	17	6	35	-3	47	2.8	37
18	16	40	18	43	7	44	18	5	33	-2	51	3.0	38
19	14	39	19	44	8	44	19	4	32	-1	62	3.2	39
20	11	38	20	45	9	46	20	2	30	0	69	3.4	40
21	9	36	21	46	10	47	21	2	29	1	75	3.6	41
22	7	35	22	46	11	48	22	1	27	2	80	3.8	42
23	5	34	23	47	12	49	23	1	24	3	84	4.0	43
24	4	33	24	48	13	50	24	18	18	4	87	4.2	43
25	3	32	25	49	14	51	25	17	17	5	90	4.4	44
26	3	31	26	50	15	52	26	13	13	6	92	4.6	45
27	2	29	27	51	16	52	≥26	6	6	7	93	4.8	46
28	1	28	28	52	17	53	53	8	8	8	94	5.0	47
29	1	27	29	52	18	54	54	9	9	9	95	5.2	48
30	1	26	30	53	19	55	55	10	10	10	96	5.4	49
31	1	24	31	54	20	56	56	11	11	11	97	5.6	49
32	1	19	32	55	21	57	57	12	12	12	97	5.8	51
≥33	1	13	33	56	22	58	58	13	13	13	98	6.0	53
			34	56	23	58	58	14	14	14	98	6.2	56
													52

Appendix B. (continued) Percentiles and T-Score Equivalents for the Experimental Control Scores

Raw Score	Control Score 1		Control Score 2		Control Score 3		Control Score 4		Control Score 5		Control Score 6	
	%	T	Raw Score	%	Raw Score	%	T	Raw Score	%	Raw Score	%	T
35	76	57	24	82	59	15	99	72	6.4	59	52	52
36	79	58	25	84	60	16	99	73	6.6	61	53	53
37	81	59	26	86	61	17	99	75	6.8	64	54	54
38	83	60	27	87	61	18	79	70	7.0	54	54	54
39	84	60	28	89	62	19	80	80	7.2	70	55	55
40	85	60	29	90	63	20	80	80	7.4	72	56	56
41	86	61	30	91	64	21	80	80	7.6	74	56	56
42	88	62	31	92	64	≥22	80	80	7.8	76	57	57
43	90	63	32	93	65	8.0	80	80	8.0	78	58	58
44	91	63	33	94	65	8.2	80	80	8.2	80	59	59
45	91	64	34	94	66	8.4	81	81	8.4	81	59	59
46	92	64	35	95	67	8.6	83	83	8.6	83	60	60
47	94	65	36	96	67	8.8	84	84	8.8	84	61	61
48	95	66	37	96	68	9.0	86	86	9.0	86	62	62
49	95	67	38	97	69	9.2	88	88	9.2	88	62	62
50	96	67	39	97	69	9.4	89	89	9.4	89	63	63
51	96	68	40	98	70	9.6	90	90	9.6	90	64	64
52	97	69	41	98	71	9.8	91	91	9.8	91	64	64
53	98	70	42	99	73	10.0	92	92	10.0	92	65	65
54	98	70	43	99	73	10.2	93	93	10.2	93	66	66
55	98	71	44	99	73	10.4	94	94	10.4	94	67	67
56	99	72	45	99	73	10.6	95	95	10.6	95	67	67
57	99	73	46	99	75	10.8	96	96	10.8	96	68	68
58	99	75	47	99	79	11.0	96	96	11.0	96	69	69
59	99	78	48	80	80	11.2	97	97	11.2	97	69	69
60	99	79	49	80	80	11.4	97	97	11.4	97	70	70
≥61	99	79	50	80	80	11.6	98	98	11.6	98	71	71
			≥51	80	80	11.8	98	98	11.8	98	72	72
						12.0	99	99	12.0	99	73	73
						12.2	99	99	12.2	99	75	75
						12.4	99	99	12.4	99	79	79
						≥13.2	80	80	≥13.2	80	80	80