

2000

Assessing potential for learning: A factor-analytic study of a performance-based identification protocol for young, socioeconomically disadvantaged high-ability learners

Robert Martin Reardon
William & Mary - School of Education

Follow this and additional works at: <https://scholarworks.wm.edu/etd>



Part of the [Educational Assessment, Evaluation, and Research Commons](#)

Recommended Citation

Reardon, Robert Martin, "Assessing potential for learning: A factor-analytic study of a performance-based identification protocol for young, socioeconomically disadvantaged high-ability learners" (2000). *Dissertations, Theses, and Masters Projects*. Paper 1550154147. <https://dx.doi.org/doi:10.25774/w4-fqyv-hp21>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Dissertations, Theses, and Masters Projects by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600**

UMI[®]

**ASSESSING POTENTIAL FOR LEARNING:
A FACTOR-ANALYTIC STUDY OF A
PERFORMANCE-BASED IDENTIFICATION PROTOCOL FOR YOUNG,
SOCIOECONOMICALLY DISADVANTAGED HIGH-ABILITY LEARNERS**

A Dissertation

Presented to

**The Faculty of the School of Education
The College of William and Mary in Virginia**

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

by

Robert Martin Reardon

May 2000

UMI Number: 9974961

UMI[®]

UMI Microform 9974961

Copyright 2000 by Bell & Howell Information and Learning Company.

All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.


Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

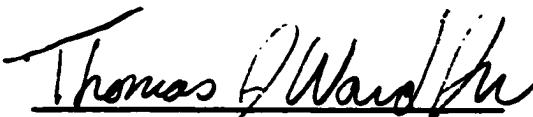
ASSESSING POTENTIAL FOR LEARNING:
A FACTOR-ANALYTIC STUDY OF A
PERFORMANCE-BASED IDENTIFICATION PROTOCOL FOR YOUNG,
SOCIOECONOMICALLY DISADVANTAGED HIGH-ABILITY LEARNERS


by

Robert Martin Reardon

Approved May 2000 by


Joyce
Chair 
Joyce L. VanTassel-Baska, Ed.D.


Thomas J. Ward, Ph.D.


Robert J. Hanny, Ph.D.

**Dedicated to my parents
Tim and Norah Reardon
for helping me to understand my potential for learning.**

**Instruction is possible only where there is potential for imitation.
(Vygotsky, 1987, pp. 210-211)**

**It is not only likely—it is inevitable—that he *make up his personality*,
under the limitations of heredity, by imitation, out of the ‘copy’
set in the actions, temper, emotions, of the persons
who build around him the social enclosure of his childhood.
(Baldwin, 1906, p. 340)**

Table of Contents

Chapter One	2
The problem	3
Equitable access	4
Disadvantaged gifted	7
Problem statement	8
Current study	9
Conceptual framework	12
Contribution of this study	16
Key Terms	17
Development	17
Learning	18
Performance-based assessment	19
.....	19
Giftedness and high-ability learning	19
Socioeconomically disadvantaged high-ability learners	21
Conclusion	23
 Chapter Two	 24
Overview	25
First Movement §1(a): Underrepresented groups	26
First Movement §1(b): The reality of socioeconomic disadvantage	29
First Movement §1(c): The role of home values	29
First Movement §1(d): Early intervention desirable	31
First Movement §2(a): Traditional methods for identifying high-ability learners	32
First Movement §2(b): Academic disagreement concerning the meaning of outcomes	34
First Movement §2(c): Appropriate tests for high-ability learners	36
The Otis-Lennon School Ability Test (OLSAT)	36

Metropolitan Achievement Test (MAT-7)	37
First Movement §2(d): “Traditional” procedures are ineffective	39
First Movement §2(e): Regulations covering gifted identification in South Carolina	40
First Movement §3(a): Alternative ways of identifying children from socioeconomically disadvantaged backgrounds	42
Portfolio identification	43
Best performance	44
Direct observation	45
Dynamic assessment	45
First Movement §3(b): Other factors to be considered	46
First Movement §4(a) The relationship between development and learning	47
First Movement §4(b) The dialectical nature of growth	51
Second Movement §5(a): Feuerstein and dynamic assessment	54
Second Movement §5(b): Another look at the zone of proximal development . .	56
Second Movement §5(c): Mediation in dynamic vs static assessment	58
Second Movement §5(d): Dynamic assessment, its derivatives, and the identification of high-ability learners	60
Second Movement §6(a): High-ability learners	61
Second Movement §6(b): Adapting dynamic assessment to fit the clientele . . .	62
Second Movement §7: Dramatism	63
Summary	67
 Chapter Three	 69
Overview	70
Research questions and instruments	70
Research Design	72
Site selection	73
Description of the test instruments	74
Basic test protocol	77
Training requirements	78

Test administrator's training	78
Test scorer's training	79
Other requirements	79
Participants	80
Other instrumentation	81
Data collection	82
Statistical Procedures	82
Time frame for this study	84
Limitations and delimitations	84
Chapter Four	88
Overview	89
Reliability	90
Summary of inquiry into reliability	102
Correlation between Project STAR and other instruments	102
Factor Analytic Questions	103
General consideration of analytical issues	104
Field test as sample.	104
Bias in general.	104
General analytical criterion.	106
Anchoring the field test sample	106
General exploratory factor analysis procedure	110
Summary of exploratory factor analyses	117
Confirmatory factor analysis	117
Grade 4	119
Grade 5	124
Grade 6	126
Findings on bias	127
Concluding remarks	129

Chapter Five	131
Introduction	132
Discussion	133
Contribution of this study	134
Remaining issues	135
The concept of dramatism	141
Conclusion	144
Implications	144
Appendix A	147
References	158

Acknowledgments

I would like to acknowledge my deep gratitude to my Dissertation Committee chair, Dr. J.L. VanTassel-Baska, and to my committee members, Dr. T.J. Ward and Dr R.J. Hanny for their encouragement, help and direction, and their enduring patience when deadlines had to be met.

Project STAR has been and still is a cooperative endeavor, and insofar as I have benefitted greatly by being able to utilize an existing database, there are a great many people to thank. From the South Carolina end of the partnership, I would like to acknowledge the indulgence of Dr. W. Lord, State Director of Gifted Education in his ready agreement to my use of the data. I must also thank Dr. Huynh Huynh of the University of South Carolina for his valued input in some of the statistics debates that arose, and particularly Donna Darby and Dr. Fran O'Tuel without whom there would be no Project STAR.

At the Center for Gifted Education, Joyce VanTassel-Baska pre-eminently deserves acknowledgment for firstly conceptualizing the instruments, and then for personally contributing the verbal items. Thanks also must go to Dana Johnson who accepted Joyce's invitation to be an integral part of the Project STAR task design committee in her then capacity as consultant to the Center for Gifted Education. Dana was ably assisted by Laura Dolins who accepted Project STAR as her graduate assistantship responsibility from the inception to the end of the pilot phase. After the successful pilot phase, Dana spent many hours preparing for the field test stage, and even since she officially relinquished her consultant role with Project STAR she has been constant in contributing ideas, revising items and editing. Thanks also to Dr. Linda Avery, the Center Manager, for her detailed organization, and ongoing intensive involvement in the teacher training, liaison, reporting, revision and documentation phases of Project STAR. Many graduate assistants at the Center for Gifted Education assisted with data input at both the pilot and field test stages, and I thank them all.

I end by returning to thank Joyce VanTassel-Baska who has gone beyond the strict requirements of a committee chair to see that I had every opportunity to complete

my doctoral degree in timely fashion, and whom I will always thank for providing the opportunity which I can now say I have truly grasped.

List of Tables

Table 1
Sampling protocol showing systematic spreading of districts across forms 74

Table 2
Correspondence between Project STAR prototypes and the South Carolina Standards 76

Table 3
Breakdown of pilot sample, showing numbers of students for whom PACT data was obtained 92

Table 4
Descriptive statistics for equating Form A results to Form B equivalents 94

Table 5
Final breakdown of pilot sample, after removal of outliers 97

Table 6
Summary of results for test of difference between B-equivalent (B*) and B scores after anchoring 99

Table 7
Coefficient alpha reliability of pilot “field test equivalent” Project STAR instruments 100

Table 8
Breakdown of field test sample, showing numbers of students for whom PACT data was obtained 106

Table 9
Descriptive statistics for equating Form A results to Form B equivalents- field test 108

Table 10
Summary of results for test of difference between B-equivalent (B*) and B scores after anchoring 110

Table 11
Two-factor solution: Factor 1 and Factor 2 for grade 4 114

Table 12
Three-factor solution: Factor 1, Factor 2, and Factor 3 for grade 5 115

Table 13
Three-factor solution: Factor 1, Factor 2, and Factor 3 for grade 6 116

Table 14
Model fit statistics bearing on the issue of bias 128

List of Figures

Figure 1 Csikszentmihalyi's (1991) depiction of the flow channel in which skill is commensurate with task difficulty. Adapted from <u>Flow: The psychology of optimal experience</u> (p. 74), by M. Csikszentmihalyi, 1991, New York: HarperCollins. Copyright 1990 by Mihaly Csikszentmihalyi.	12
Figure 2 A depiction of the zone of proximal development showing varying widths of the zone for different tasks—in the style of Csikszentmihalyi (1991).	14
Figure 3 Graphic organizer for the literature review	25
Figure 4 Types of standardized tests in American schools. (USGAO, 1993, p.21)	33
Figure 5 Grade 4 Intermediate Form A group STAR regressed on PACT, showing outliers on PACT results.	96
Figure 6 Grade 4 Intermediate Form A STAR regressed on PACT, showing the effect of removing the three PACT outliers in Fig. 5.	96
Figure 7 Scree plot for grade 4 exploratory data using principal component analysis: showing the two eigenvalues in the steep part of the graph.	113
Figure 8 Structural equation model for a priori assignment of items to domains in the design process for the grade 4 Project STAR instrument.	119
Figure 9 Two-factor solution for Grade 4, arising out of the exploratory factor analysis.	122
Figure 10 "Best-fit" structural equation model for grade 4 data based on exploratory factor analysis.	123
Figure 11 Structure of the priori model at grade 5 level.	124
Figure 12 Three-factor model for the grade 5 data reflected the outcome of the exploratory factor analysis.	125
Figure 13 The 1-factor solution for the grade 5 data from the exploratory factor analysis.	126

**ASSESSING POTENTIAL FOR LEARNING:
A FACTOR-ANALYTIC STUDY OF A
PERFORMANCE-BASED IDENTIFICATION PROTOCOL FOR YOUNG,
SOCIOECONOMICALLY DISADVANTAGED HIGH-ABILITY LEARNERS**

ABSTRACT

This factor-analytic study of a performance-based identification protocol for young, socioeconomically disadvantaged high-ability learners investigated the issues of reliability, test equivalency, and bias. A group-administered, performance-based set of instruments was designed in a joint project between the Center for Gifted Education and the State Department of Education, South Carolina. These instruments went through a series of processes of review and refinement leading to their use in a field test in fall 1999. The outcome of this field test administration is the subject of exploratory and confirmatory factor analysis in this study.

Reliability of the instruments was established on the pilot study data which were gathered from a heterogeneous sample of 1425 students. Statistical anchoring using linear transforms was used to address the status of the two forms of the test instruments. The Cronbach alpha values ranged from 0.71 to 0.78, values lower than desirable for psychometric instruments, but acceptable in view of the special purpose of this test. Exploratory factor analysis on a randomly chosen half of the field test data (N = 1800 students) lead to structural equation modeling of both a priori and exploratory factors on the second half of the field test data.

The exploratory factor analyses did not support a construct of high-ability learning. All emergent factors accounted for less than a majority of the variance in the relevant sub-samples. Nonetheless, the structural equation models demonstrated that there was no evidence of bias on the basis of gender, ethnicity, or socioeconomic status. Project STAR did indeed exhibit the ability to discriminate in an unbiased way among young, socioeconomically disadvantaged high-ability learners.

The overriding implication of this study is that performance-based identification should be utilized as part of the testing battery available to school districts seeking to assess potential for learning. At the same time, the failure to detect a strong factorial

structure in the results of a performance-based test specifically designed around a factorial schema implies that there are layers of complexity inherent in this testing protocol that deserve close attention. Further research arising from increasingly standardized implementations is expected to shed more light on what has been called in this study the “elusive factor” issue.

**ROBERT MARTIN REARDON
EDUCATIONAL POLICY, PLANNING, AND LEADERSHIP (GIFTED EDUCATION)
THE COLLEGE OF WILLIAM AND MARY IN VIRGINIA**

**ASSESSING POTENTIAL FOR LEARNING:
A FACTOR-ANALYTIC STUDY OF A
PERFORMANCE-BASED IDENTIFICATION PROTOCOL FOR YOUNG,
SOCIOECONOMICALLY DISADVANTAGED HIGH-ABILITY LEARNERS**

Chapter One

The Problem

The problem

To be educated is to aspire to one's potential as a human being. To control the access to advanced education is therefore to exercise enormous power, and such power brings with it the responsibility to be accountable for one's stewardship. History attests to the institutionalized restriction by gatekeepers, albeit often acting inadvertently, of the access to higher education for major sections of humanity. Women, for instance, have for many years suffered from an inability to claim even their own place in history: "Men have had every advantage of us in telling their own story. Education has been theirs in so much higher a degree; the pen has been in their hands" (Austen, 1818).

The hand holding the pen is a worthy icon for this study in which the key issue concerns placing the pen in the hand best qualified to wield it, regardless of accrued advantage from factors like socioeconomic status. Frasier (1993) declared that "a universal problem in the field of gifted education is the identification and nurturance of talented students from disadvantaged and culturally different backgrounds" (p. 685). The National Excellence report (OERI, 1993) found that the economically disadvantaged were significantly underserved, with "only 9 percent of students in gifted and talented education programs...in the bottom quartile of family income, while 47 percent of program participants were from the top quartile in family income" (p. 17). At the heart of this study is the proposition that if one is controlling access to appropriate educational opportunities for high-ability learners, to use demonstrated educational prowess as the only access criterion is to inherently bias the selection process. To use demonstrated aptitude or achievement as sole criteria means that those who have been hampered in, or even disqualified from antecedent educational opportunities by accidents of fate are going to be unfairly deprived. Further, should such deprived people be socioeconomically

disadvantaged they are all the less able to know how to be heard in their complaints. To use sound, unbiased techniques to control access to higher, more advanced education, on the other hand, is not only logical, it is just.

Equitable access

At the federal level, concern over the low levels of overall achievement of high school graduates has not abated since A Nation at Risk (1983), and state legislatures are similarly rightly concerned, and thus have instituted standards of quality or similar expectations for the outcomes of schooling. These have given rise to standards of accreditation which school districts and individual schools must meet in order to remain in operation. Reform of current practice is demanded at all levels.

In the face of the worthy motivation of standards-based reform-- “high standards for all students” (McLaughlin & Shepard, 1995, p. xvi)--and the resulting emphasis on equality of outcome, not just opportunity, it is salutary to ponder the fate of “even the best American students (who) do not fare well in international comparisons” (McLaughlin & Shepard, 1995, p. 1). It is unfortunate that the decision process at the local level can come down to an “either...or” choice in terms of striving for equality of outcome juxtaposed against developing the talents of each child to the fullest extent.

Equitable access to the services provided in the school district for high-ability learners is a more realistic goal than “high standards for all students,” but it has proven to be exceedingly elusive. Statistics continue to show an imbalance along ethnic lines. The 1988 National Education Longitudinal Study reported that 8.8% of all 8th-grade students in public schools participated in gifted and talented programs, according to the National Excellence report (OERI, 1993, p. 17). Of these, racial and ethnic groups were represented as follows: 17.6% of Asian students, 9.0% of white, non-Hispanic students,

7.9% of black students, 6.7% of Hispanic students, and 2.1% of American Indian students (OERI, 1993, p. 17).

Several categories of talented students are particularly neglected in programs for top students. These include culturally different children (including minority and economically disadvantaged students), females (who are underserved in mathematics and science programs), students with disabilities, high potential students who underachieve in school, and students with artistic talent. (OERI, 1993, p.16)

National Excellence (OERI, 1993) is not alone in drawing attention to the plight of the socioeconomically disadvantaged. Writing about those of disadvantaged and culturally different backgrounds, Frasier (1993) commented that “children from either group are disproportionately underrepresented in talent development programs” (p. 685). Borland, Schnur and Wright (2000) have recently commented that “the underrepresentation of economically disadvantaged children and adolescents—especially from racial and ethnic minority groups—in programs for gifted students is one of the most recalcitrant and troubling issues confronting educators of gifted students” (p. 13). This reinforces Borland and Wright’s earlier assertions to similar effect (1994), and is in accord with comments by Kearney and LeBlanc (1993), Passow (1989), Richert (1987), and VanTassel-Baska, Patton and Prillaman, (1989). While it is certainly undesirable, the imbalances noted by all the above authors are not unexpected in view of the heavy reliance on IQ-related identification procedures, and teacher recommendations. The considerable racial imbalance in terms of measured IQ between majority and minority populations, for example, is a well-known phenomenon, for all its unsettling overtones (Carroll, 1997; Gottfredson, 1997).

In recent times, The College Board (1999) noted in a report on minority high-achievement that “underrepresented minority students accounted for only about 1 in 20 of the students in 1998 who had the very high SAT 1 scores typical of individuals admitted to highly selective colleges and universities” (p. 7). The report pointed out that the National Assessment of Educational Progress (NAEP) data showed that substantial test score gains were made by underrepresented minorities over the last thirty years in some areas, notably in reading and mathematics, to the stage where the gap in NAEP math scores in the mid-1990s was only about a third what it had been in the early 1970s. This encouraging trend was balanced by the report’s judgement that “in some instances, ground may actually have been lost relative to Whites” (p. 6). The report went on to point out that “the large achievement gaps that persist among groups emerge very early in the students’ school careers. Indeed, national studies have found that underrepresented minorities are not performing nearly as well as White students early in the first grade and that the very large gaps identified by NAEP develop rapidly during the first three years of school” (p. 7).

It is not the contention of this study that students identified by IQ tests should be refused access to programs designed specifically for high-ability learners. Indeed, as Gagne (1997) has pointed out, the fact that IQ is “use(d) in most school districts and in most empirical studies for identification purposes, whether in the U.S., or in Canada or any other country, confirm(s) that a high IQ is THE operational definition of giftedness” (p. 78). High IQ students are patently able to benefit from special programs. However, it is the contention of this study that the use of a multi-phase identification process is essential if potential for learning at a high-level is to be determined, especially among socioeconomically disadvantaged populations. Measured IQ is clearly a relevant indicator

of ability level, but, as will be discussed at some length later, it is by no means clear to what degree one can defend the often made assumption that it alone is the only relevant measure of potential. Rowe (1997) critiques what he characterizes as the exposure theory of intelligence, namely “that intelligence is the sum total of learning experiences to which individuals have been exposed” (p. 134), by pointing out that “children will gain unequally with each exposure” (p. 135): a point made in great detail by Carroll (1997). It is the likelihood of unequal gain, while taking into account the strong arguments of Plomin and Petrill (1997) for the heritability of IQ, that supports the contention that identification should not rest solely on IQ.

Disadvantaged gifted

All reasonable efforts must be made to avoid the possible wastage of talent should high-ability learners be denied the opportunity to develop their potential—should the capable hand be denied access to the pen. In a singular contribution on the subject, Csikszentmihalyi, Rathunde, and Whalen (1993) lamented the documented waste of talent in fields as different as athletics, art, science, mathematics, mathematics and science, and music. Passow and Frasier (1996) traced concern about the wastage of talent among the underserved populations on the federal front from Education of the Gifted, produced by the Educational Policies Commission (1950), through the “Marland Report” (1971), and the “Marland Definition” (Public Law 91-230, section 806) in 1972, to the Jacob K. Javits Gifted and Talented Education Act of 1988 (Public Law 100-97). “The Javits Act reaffirmed that in every population there are individuals with potential for superior or outstanding achievement who are in environments where this aptitude may not be recognized or nurtured. These individuals are most likely to come from racial/ethnic minority or economically disadvantaged groups” (Passow and Frasier, 1996,

p.198).

It is to be expected that inherently fair identification procedures (namely, procedures that do not penalize a child for not knowing what he or she has had no chance to learn), hold promise for correcting the underrepresentation of those “from racial/ethnic minority or economically disadvantaged groups” (Passow & Frasier, 1996, p. 8) among the ranks of high-ability learners.

Problem statement

Students who can be assessed as high-ability learners (by which is meant that they can learn relatively novel skills at a single exposure) should be included in programs designed to cater for such students even if they rank somewhat below the usual percentile cut-off scores on traditional ability and achievement measures. A technique for estimating the learning potential of such students exists. Although this technique was initially developed for a population distinctly different from this proposed use, the principles invoked by the technique may be adaptable enough to implement in a group and powerful enough to be defensible as an identification methodology. The implementation of performance-based identification protocols in a dynamic-like assessment environment may enable different inflections to be added to the identification process and promise a more effective outcome to the good of all.

To this end, this study investigates whether an instance of a performance-based identification protocol can be used to select individuals from among a sample of young, high-scorers on traditional instruments, who nevertheless do not qualify for services as being academically gifted. There are a number of specific properties which it is hoped this instrument will display: it will be psychometrically robust, spread a tightly defined selection of high scoring children along a new axis called potential for learning, identify

children regardless of gender, identify children regardless of ethnicity, and most importantly, identify children regardless of socioeconomic disadvantage.

The phrase “regardless of” is used to signify that it is hoped that there will be no statistically significant difference between the groups when the children are grouped according to gender, ethnicity, and most importantly, socioeconomic disadvantage.

Two forms of the instrument will be tested in this study, yielding the ability to compare and contrast the outcomes, as well as pointing to the replicability of the protocol.

Current study

This study examined data gathered during both the pilot stage (spring semester 1998-99) and the field test stage (fall semester of 1999-2000 school year) of a suite of performance-based identification instruments developed at the Center for Gifted Education, The College of William and Mary, Virginia, USA. The instruments were developed for the State Department of Education, South Carolina as Phase I of Project STAR (Student Task Assessments and Rubrics). One thrust of Project STAR was to investigate the extent to which performance-based identification would assist in achieving a more ethnically balanced clientele for South Carolina’s programs for high-ability learners. The specific thrust of this research study took the analysis performed as part of the reporting for Project STAR itself and extrapolated it. Project STAR reported on the use of this suite of instruments in differentiating among a sample of students who were high on traditional measures of either aptitude or achievement, but not both, and who were classifiable as being from socioeconomically disadvantaged backgrounds. Of particular interest was the extent to which those from socioeconomically disadvantaged backgrounds were able to keep pace with the performance of their grade level peers. The

underlying assumption of the assessment task construction was that problem-based identification would differentiate among students functioning at a high level.

Observations made by those concerned with special programs for high-ability learners in South Carolina in recent years revealed that some ethnic groups were underrepresented. For example, those of African-American descent constitute 34 percent of the South Carolina population, but only 14 percent of the students in the high-ability learner programs are African-American. (Darby, 2000) As noted earlier, National Excellence (OERI, 1993) directed attention to the underserving of the economically disadvantaged in programs for the gifted. The correlation between ethnicity and socioeconomic disadvantage will be addressed later in this study.

South Carolina is not alone in exhibiting such phenomena. Gallagher (1998) commented somewhat wryly on the involvement of the Office of Civil Rights in querying the field of gifted education as a whole as to whether “some disguised resegregation process (was) at work” (p. 10). South Carolina does not use a single identification instrument to select those eligible for special programs, but rather considers ranks on measures of aptitude and measures of achievement. However, to be included among those receiving specially designed programs, a child must perform at very high levels on both an aptitude and achievement measure. (The details of the identification procedure are included in Chapter 2.) Under this identification policy, imbalance in the ethnic makeup of the group of identified children has been relatively invariant over recent years (Darby, 2000). A large number of ethnically diverse children qualify on, for example, aptitude, but not on achievement, or vice versa. The ethnic imbalance in South Carolina’s programs for high-ability learners exemplifies the imbalance characteristic of many programs for such children, and drew the attention of the Office of Civil Rights in 1998

(Darby, 2000).

As stated earlier, the performance of the group of near-identified students, i.e. those reaching the cut-off on one or other of either ability or achievement measures but not both, is the focus of this study. Within this group, the subset of socioeconomically disadvantaged children can be identified by their eligibility for free/reduced lunch status, and the performance of this group was compared and contrasted with the performance of those not identified as socioeconomically disadvantaged. Differences were explored across all measures at the level of domain. The domains are verbal (for both intermediate and primary), mathematical and spatial (for intermediate), and non-verbal (a composite of mathematical and spatial) for primary. Item level analysis was also pursued. The issue of the makeup of the socioeconomically disadvantaged group in terms of ethnicity and gender was also examined.

Factor analysis of the responses of the socioeconomically disadvantaged group was considered in relation to the responses of the group not identified as socioeconomically disadvantaged. The intention was to investigate whether the same factor structure appeared to underlie the responses of both groups. The results on the Otis-Lennon School Ability Test (OLSAT), the Metropolitan Achievement Test (MAT-7), and the South Carolina PACT test assisted in shedding light on group differences in as much as they provide a way of establishing how different the socioeconomically disadvantaged subset was from those not identified as socioeconomically disadvantaged prior to the administration of the performance-based instrument.

Conceptual framework

Vygotsky's (1978) concept of the zone of proximal development (Figure 1) is proposed as an adequate theoretical lens through which to view the results. A different and potentially deeper perspective on the phenomenon of performance-based identification in South Carolina was synthesized by reference to Burke's (1966, 1969a, 1969b, 1972, 1984) multiple perspectives on human action and motivation (Wertsch, 1998).

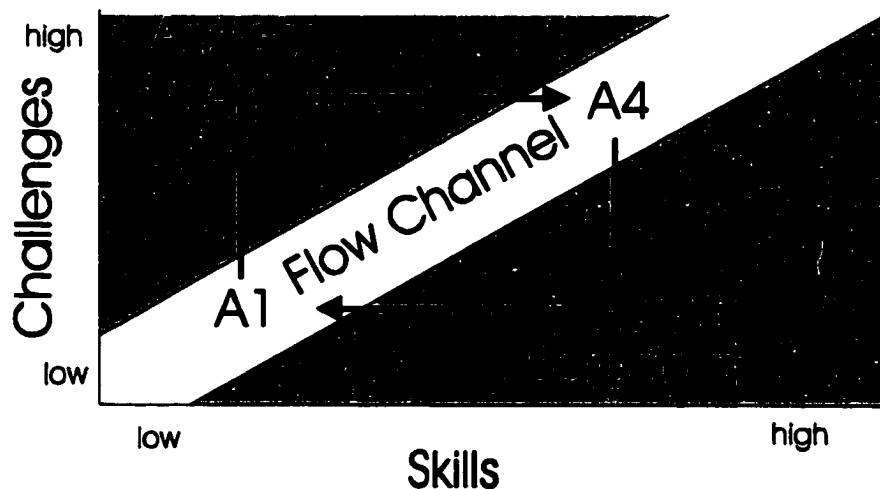


Figure 1 Csikszentmihalyi's (1991) depiction of the flow channel in which skill is commensurate with task difficulty. Adapted from Flow: The psychology of optimal experience (p. 74), by M. Csikszentmihalyi, 1991, New York: HarperCollins. Copyright 1990 by Mihaly Csikszentmihalyi.

The zone of proximal development was a central part of Vygotsky's (1978) two-fold answer to the issue of the relationship between learning and development. He subdivided this into two sub-issues: "first, the general relation between learning and development; and second, the specific features of this relationship when children reach school age" (p. 84). Of fundamental significance to Vygotsky was "the notion that what children can do with the assistance of others might be in some sense even more indicative of their mental development than what they can do alone." (p. 85). Vygotsky's formal

definition of the zone of proximal development was “the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers” (p. 86). Vygotsky was seeking to define “those functions that have not yet matured but are in the process of maturation, functions that will mature tomorrow but are currently in an embryonic state” (p. 86).

In his study of the phenomenon of “flow,” Csikszentmihalyi (1991) portrayed his explanation of why the complexity of consciousness increased as a result of optimal experiences (Figure 1). Csikszentmihalyi was interested in depicting the shift from absorption in a task at A1 (A1 being anywhere in the flow channel), to anxiety in the task at A3 where the challenge is too far in excess of skill, to absorption again at a different (higher) skill level at A4, to boredom at A2 arising from a lack of challenge in an area of high skill for the individual. Vygotsky’s (1978) antecedent zone of proximal development can be re-formulated diagrammatically in terms of Csikszentmihalyi’s (1991) flow channel. In this understanding, challenges from outside one’s zone of proximal development would not elicit engagement or produce learning, but challenges from within the central white “zone” would. Following the essence of Vygotsky’s (1978) thought, one’s “zone” would not necessarily be of uniform width in all areas of learning.¹ Figure 2 offers a visual interpretation of this understanding of the zone of proximal development for one student.

1

The width of the flow channel was not essentially uniform in Csikszentmihalyi’s (1991) thought either, although his figure did indicate uniformity, as shown in Figure 1.

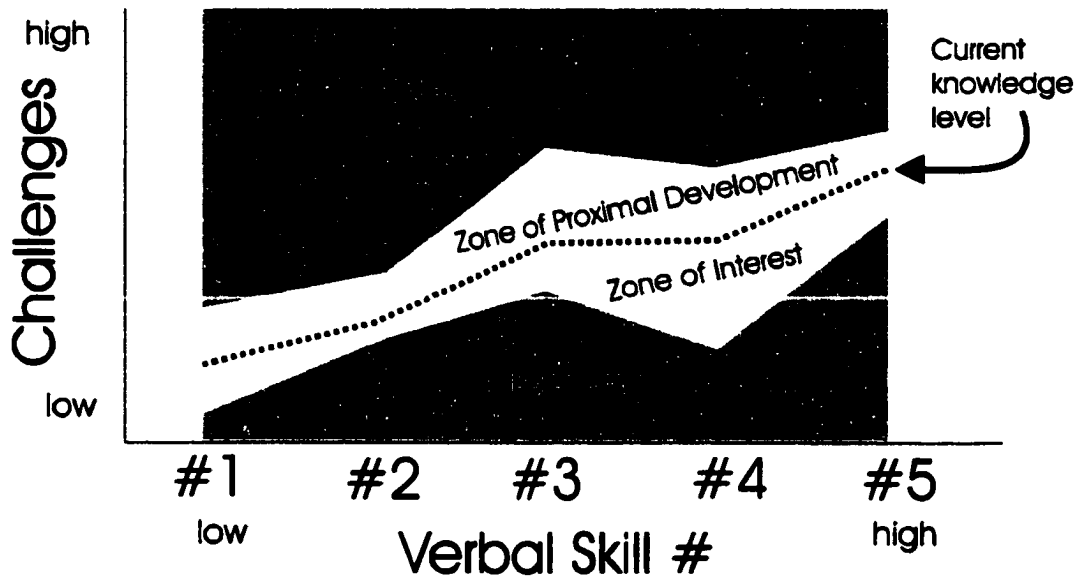


Figure 2 A depiction of the zone of proximal development showing varying widths of the zone for different tasks—in the style of Csikszentmihalyi (1991).

One of these visuals would have to be drawn for each of the domains involved in this study (verbal and nonverbal), for example, and, in practice, for all other domains of learning. The child's current level of learning is depicted as the main broken line with the zone of proximal development extending above it until it merges into the realm of anxiety. Below the child's current level of learning in Figure 1 is an area labeled the zone of interest, corresponding to those tasks which the child finds motivating, even though he or she is already adept in them. Below this zone of interest is the realm of boredom, where the child can already perform the task, and where the task no longer holds any interest for him or her.

One clear implication arising from this discussion is that if a given task is within the individual's zone of proximal development, and if the motivation is appropriate, the individual will learn at an optimal rate. In Figure 2, while the zones of proximal

development and interest are drawn continuously, the tasks themselves are depicted as points along the continuum from low to high. Tasks which involve challenges from outside the individual's zone of proximal development will engender either boredom (too far below one's skill level to be still interesting), or disinterest for the sake of self-image preservation (too far above one's skill level). A further implication relevant to this study is that by demonstrating a difficult task (preteaching), perhaps it can be made comprehensible to those whose zone is amenable to the task—even differentiating among those students who may appear to be “in the same neighborhood” on the basis of some other criterion.

In dynamic assessment, the child is first shown how to perform a task which presents challenges beyond his or her zone of proximal development. After having been exposed to the preteaching to the stage where the challenge level is lowered, the child is then asked to perform a similar task. If the child can perform the task, it is an indication of the malleability and depth of the zone of proximal development of the child; more an indication of the child's potential for learning than a static declaration of attainment in learning. As Vygotsky (1978) further remarked: “the zone of proximal development permits us to delineate the child's immediate future and his dynamic developmental state, allowing not only for what already has been achieved developmentally but also for what is in the course of maturing.” (p. 87). The promise of the zone of proximal development concept in providing a perspective into the immediate future, drawing on information concerning a student's dynamic developmental state, is its allure in this context.

It is in this vein that Burke's (1968) conceptualization of dramatism seems to provide a milieu in which to consider specifically the dynamic nature of Vygotsky's (1978) zone of proximal development. Dramatism “takes human action as the basic

phenomenon to be analyzed” (Wertsch, 1998, p. 12; emphasis in the original). Burke’s understanding of action was inseparable from the motivation for the action; motivation is “what is involved when we say what people are doing and why they are doing it” (1969a, p. xv). Wertsch (1998) points out the clear linkage between this dramatic method and the work of Vygotsky, among others, in that it takes “human action to be (the) fundamental unit of analysis” (p. 12).

Burke’s (1968) insights provide an interpretive lens through which to view the wealth of detail contained in Project STAR—a way of stepping back to consider what is really at issue, namely, the inclusion of socioeconomically disadvantaged, high-ability learners in programs which will be to their advantage.

Contribution of this study

What is being sought through Project STAR is evidence that there is some propensity being tapped by the Project STAR instruments, identifiable as high-ability learning potential, that correlates with accepted standardized measures of high aptitude and/or achievement, but which is different in that it adds another dimension to the identification process. The importance of seeking to use pertinent identification instruments targeted specifically at the high-ability learner was cogently underlined by Sternberg and Zhang (1995) when they commented, concerning constructs or measures that should be used to identify the gifted: “If we care about the potential of an individual to contribute to him/herself, others, and society in a productive way, then we need to justify why the measures we use will help identify such potentially productive individuals” (p. 93).

If the existence of high-ability learning as a characteristic related to but different from those detected by traditional aptitude and achievement measures, can be

demonstrated by the use of performance-based instruments, then the option to use these instruments in conjunction with the more usual measures is very worthwhile. Astute use of such instruments will go a long way towards addressing the concern held by many, and so well expressed by Gallagher (1998), that “by the time these minority gifted students reach us in the upper grades they are truly behind” (p.11). By clearly establishing a defensible procedure to enable socioeconomically disadvantaged students to be identified, this study hopes to designate areas of enhanced return for educational testing effort invested. In other words, it is hoped that the use of performance-based identification will identify a group of previously unidentified children who will benefit from, and are deserving of, purposeful program adaptations for high-ability learners. Thus, this study is expected to yield valuable insight into the potential of such performance-based identification for redressing socioeconomic imbalances within the set of identified high-ability learners, thereby contributing to the placing of pens in hands well suited to wield them.

Key Terms

Development

The term “development” in this study is contrasted with “learning.” From a dynamic system model approach, Thelen and Smith (1994) declared that “order, discontinuities, and new forms emerge precisely from the complex interactions of many heterogeneous forces” (p.37). Schooling provides one environment in which these complex interactions occur. Expanding on this, van Geert (1998) declared that “developmental order comes about as a result of self-organization” (p. 635). While the concept of order arising spontaneously out of complexity seems counterintuitive, Elman et al. (1996) have developed a neural network model in which a specified input, for

instance, of natural language in a communicative context, gave rise to a targeted output of, for instance, a child's ability to talk in sentences that comply with the grammar of his or her mother tongue. In this instance, the proposed mechanism intermediating between the input and output is a self-organizing structure of interconnected nodes—an analog of a brain in a particular stage of development.

This research study involved students who would be adjudged by their grade level and their age to be at approximately the same developmental level. Students who were in grades 3 and 4 were classified as being in the primary level group, and those who were in grades 5 and 6 were grouped in the intermediate level. The only implication is that these children are presumed to possess roughly commensurate school acculturation. For example, it is presumed that children can read to some degree, can write sufficiently well to be legible to the teachers who will score the response booklets, can be expected to know the usual protocol for asking questions in class, and so forth. The expectation is that children in the two forms will be roughly at the same developmental level—an assumption which will be tested in due course. One of the strengths of the modified dynamic assessment approach implemented here is precisely that such minimal developmental assumptions regarding content knowledge have to be made.

Learning

Learning connotes a conscious engagement with an environment. For Piaget (1970), a major explanatory mechanism of learning was what he referred to as “adaptation.” This general process applied to biological interactions as well as to the realm of cognitive operations, and resulted from the interaction of two opposing tendencies: assimilation and accommodation. Assimilation was “the integration of external elements into evolving or completed structures of an organism” (Piaget, 1970,

pp. 706-707). It corresponded, for example, to a child's attempts to understand an experience of the world in terms of already established cognitive understandings.

In this research study, learning is inferred when the students participating in the study successfully perform a task after suitable pre-teaching. No conclusions are drawn about the student's adaptive or assimilative changes, although it is reasonable to assume that such processes may have been involved.

Performance-based assessment

The concept of performance-based assessment connotes the intention to "engage students in 'real world' tasks rather than multiple choice tests, and evaluate them according to criteria that are important for actual performance in that field" (Darling-Hammond, 1995; cf Wiggins, 1990). In terms of Project STAR, this signified that a number of items emphasized concrete referents, while for others it meant that the students were actually encouraged to develop solutions, using some manipulatives like counters or beans. By performing certain actions and recording the outcomes of certain tasks the student can demonstrate that he/she has the ability to comprehend the rationale behind certain tasks. This enables learning potential to be inferred.

The concept of assessing what a student knows by verifying what he or she can do is both intuitively satisfying and deceptively simple. Putting the idea into practice, once one moves outside a strictly practical field like bricklaying, for example, can be problematic because the performance of a relatively culture-free action might not imply the ability to perform many other possibly culture-loaded actions, like the ability to write in English, with which formal educational establishments are vitally concerned.

Giftedness and high-ability learning

The term high-ability learning was preferred in this study to the more usual term

giftedness, because of the specific focus of the study. In choosing a distinct term, it was not the intention to suggest that gifted and talented students were not high-ability learners. In the context of this study, however, students who were not gifted by virtue of state definition were being identified as eligible to receive the attention allocated to gifted children. Hence it was thought helpful to maintain the use of a separate term to refer to these children. However it was not the primary intention to discriminate among the gifted on the basis of some being high-ability learners while others were not.

Depending on the outcome of the ensuing factorial analysis, a high-ability learner will be defined as one who can perform at a consistently high level on a number of challenging, relatively novel tasks on which he or she has been given specific instruction immediately prior to the task. Identification as academically gifted was understood in this study as it was defined in the revised South Carolina State Regulations. Lee and Lord (1999) set out the criteria in two complementary sections of the same document as follows:

The following students are deemed eligible for services with the approval of the District Evaluation Placement Team:

...

b. Students who meet the criteria in two out of three dimensions that follow.

c. Students who meet the 96th national age percentile composite score or higher (placement grades 3-12) or the 98th national age percentile composite score or higher (placement grades 1-2) on an individual or group aptitude test.

(p. 5-6)

The second of the two sections provides that in Dimension A students will be at the 90th or greater national age percentile in “one or more of verbal/linguistic,

quantitative/mathematical, non-verbal, and/or a composite of the three” (Lee & Lord, 1999, p. 7). Dimension B stipulates that students will be at or above the 94th national percentile “in reading and/or mathematical areas as measured by nationally normed or South Carolina statewide assessment instruments” (Lee & Lord, 1999, p. 7).

High-ability learning in the spirit of Dimension C of the South Carolina State Regulations for screening/referral/assessment is defined in terms of a set of characteristics. “Characteristics for this dimension are demonstrated through: ... (b) Assessments of performance tasks for placement in Grades 1-6...The performance standard is four points on a five point scale” (Lee & Lord, 1999, p. 7-8). The South Carolina Regulations will be dealt with more fully in Chapter 2.

Socioeconomically disadvantaged high-ability learners

VanTassel-Baska (1991) noted a lack of consensus concerning the definition of the term “disadvantaged,” with some studies concentrating on the reality of economic hardship (VanTassel-Baska & Willis, 1987), while others concentrated on the father’s educational and occupational status (Jencks, 1972), or minority status and cultural values per se (Baldwin, 1985; Frasier, 1980). VanTassel-Baska, Patton, and Prillaman (1989) suggested that the term “disadvantaged” itself should be abandoned in favor of “at-risk for accessing educational advantages in the larger society.”

In this study a socioeconomically disadvantaged student was defined as one who was identified by the school district as being eligible for a free or reduced lunch. The free or reduced lunch program is maintained by the Food and Nutrition Service, United States Department of Agriculture. Annual adjustments are published each year in a set of guidelines “intended to direct benefits to those children most in need and are revised annually to account for changes in the Consumer Price Index” (FNS, USDA, 1999, p.

15951). Schools and institutions which charge for meals separately from other fees are required to “serve free meals to all children from any household with income at or below 130 percent of the poverty guidelines...serve reduced price meals to all children from any household with income higher than 130 percent of the poverty guidelines, but at or below 185 percent of the poverty guidelines” (FNS, USDA, 1999, p. 1591).

While the poverty thresholds are “the original version of the federal poverty measure” (ASPE, USHHS, 2000, p. 7555), and are used for statistical purposes, the poverty guidelines are “a simplification of the poverty thresholds for use for administrative purposes—for instance, determining financial eligibility for certain federal programs” (ASPE, HHS, 2000, p. 7555). The guidelines stipulate cut-off income levels for family units from size 1 through 8, with a uniform increment for each additional person above 8, for the 48 contiguous states, Alaska and Hawaii. In terms of school-based research, the free/reduced lunch students are often grouped into one set of students (e.g. National Center for Educational Statistics, USDOE, 1998, p. 22, 25, 26 ff.), and this precedent is followed in this study. Within this group of students there is clearly room for much greater differentiation in terms of parent income, but such inquiries into the parental income details would have been difficult to institute and could well have been viewed as unwarranted intrusion.

Defining socioeconomic disadvantage as equivalent to eligibility for free/reduced lunch, while potentially covering quite a range of household financial circumstances, provides a readily accessible and non-intrusive way of delineating a group of students who could be expected to be at some disadvantage, at least, when compared with their more well-off peers in taking traditional ability and achievement tests (Natriello, McDill, & Pallas, 1990; Federal Interagency Forum on Child and Family Statistics, 1998).

Conclusion

To ignore the obvious imbalance in the ethnic breakdown of those identified as high-ability learners is to run the risk of perpetuating injustice. Ethnicity is frequently entangled with socioeconomic disadvantage. If the ethnic imbalance is in fact due to the use of inappropriate or biased identification instruments and procedures, steps to redress the imbalance are clearly demanded. There is some evidence to indicate that identifying students by concentrating more directly on their ability to learn, rather than focusing on static measures which are indirect indicators of learning potential, may help to correct the imbalance. If the instrument specially designed for this study fulfills its promise, the ranks of those identified as academically gifted will be swelled by a new cadre of students—high-ability learners—identified in an unbiased way regardless of gender, ethnicity or socioeconomic disadvantage. This would represent a major step in validating the worthiness of the hand ultimately holding the pen.

Chapter Two

Literature Review

Overview

The idea that one should be able to identify someone with high-ability to learn by engaging him or her in a learning task and observing how he or she performs is intuitively satisfying. Linn and Niemi (1995) declare that “the logic of student performance assessment is compelling” (p. 197). This study is concerned with using performance as a basis for identification of high-ability learners, paying special attention to the effects on those children from socioeconomically disadvantaged backgrounds. A high-ability learner in this study is a child who can perform at a consistently high level on a number of challenging, relatively novel tasks on which he or she has been given specific instruction immediately prior to the task.

This chapter is subdivided into a total of seven sections, which have been arranged to form two movements as in Figure 3. The first movement, corresponding to



Figure 3 Graphic organizer for the literature review

the top three points of the star and the top arch, consists of a total of four sections, three of which which build systematically to the point where it becomes important to pause and consider the fourth section--one of the two concepts spanning this study--what is the relationship between development and learning. These first three sections open with a summary overview of seminal literature concerning the current topography of the field on the issue of socioeconomically disadvantaged children and gifted education programs. The second section focuses on key aspects of identification as it has been practiced, and sets about sewing seeds of discontent with "business as usual." The third section then presents some alternative identification procedures, lightly prefiguring the later treatment of dynamic assessment.

The second movement commences with a more in-depth look at dynamic assessment. This gives way to a section which blends the pertinent characteristics of high-ability learners with the insights developed into dynamic assessment to conclude with proposing performance-based identification as a modification of dynamic assessment appropriate to group administration scenarios. The second movement closes with a consideration of dramatism--the second of the two overarching concepts spanning this study.

First Movement §1(a): Underrepresented groups

The identification of individuals who will receive the benefits of whatever services the particular school district deems appropriate for high-ability learners has been an intractable problem for many years (Borland, 1989; Borland & Wright, 1994; Gallagher, 1985; Howley, Howley & Pendarvis, 1986; Pendarvis & Howley, 1996; Tannenbaum, 1983). This problem is even more pronounced when one examines the proportion of high-ability learners selected from among diverse and culturally different

groups. Richert (1985) nominated certain groups of students as being consistently underrepresented in gifted programs, namely:

(a) underachieving, poor and minority gifted children who most need programs to develop their potential; (b) the creative and/or divergent thinkers whose abilities are not tested by standardized intelligence or achievement tests or grades; and (c) other groups including the learning disabled or handicapped gifted. (p. 70)

Frasier's (1987) findings concurred with Richert's, and the situation does not appear to have changed significantly in the ensuing years. Baldwin (1991) made essentially the same observation, and Hunsaker (1994), in agreeing, based his conclusions on a survey of 56 rural or urban sites in which the median income was below the poverty line and where at least 5% of the population were minority students.

To draw attention to the persistence of the problem of underrepresentation, particularly from among the ranks of the socioeconomically disadvantaged, is not to denigrate the many remedial efforts which have been mounted, and the close attention that some school districts have bestowed on the problem. For example, Charlotte-Mecklenberg Public Schools have long wrestled with problems of racial balance in schools and programs, beginning with the landmark Supreme Court case *Swann v. Charlotte-Mecklenberg Board of Education* (1971) which established busing as one of a number of judicially acceptable alternatives to remedy de jure segregation. In 1993 Charlotte-Mecklenberg Public Schools reported that: "One of the great difficulties in American education has been the emergence of special programs for the academically gifted, programs that are dependent on or limited by assessment instruments that do not represent the fairest means for selecting students from diverse and culturally different groups" (1993, p.6). The Charlotte-Mecklenberg twofold response to the problem

embraced instituting high quality instructional programs beginning in kindergarten, and implementing strategies to meaningfully integrate academically gifted programs. A major strategy for achieving meaningful integration was to institute identification procedures which were perceived as being fair.

If one accepts the proposition that the proportion of high-ability learners should be roughly the same across all population subsections (Frasier, 1987; Hunsaker, 1994), then there is every reason for those engaged in identification to be concerned by the solid evidence that children from certain subsections are being disproportionately under-identified. Hunsaker (1994) pointed out that despite two decades of research and demonstration projects, “among the most troubling issues in education is the persistent underrepresentation of nonwhite, economically disadvantaged populations in gifted programs” (p.72). It is pertinent to draw attention to the juxtaposition of the “nonwhite” and “economically disadvantaged” in Hunsaker’s declaration.

The National Excellence report (OERI, 1993) declared that “schools must eliminate barriers to participation of economically disadvantaged and minority students in services for students with outstanding talents (and) must develop strategies to serve students from underrepresented groups” (p.28). National Excellence relied on the 1988 National Education Longitudinal Study (NELS) to assert that some minority groups were more likely to be served than others. For instance, the 1988 NELS study found that about 8.8% of all 8th-grade public school students participated in gifted and talented programs. Racial and ethnic groups were represented as follows: 17.6 % of Asian students; 9.0% of white, non-Hispanic students; 7.9% of black students; 6.7% of Hispanic students; and 2.1% of American Indian students.

First Movement §1(b): The reality of socioeconomic disadvantage

As VanTassel-Baska (1991) pointed out, socioeconomic disadvantage is frequently entangled with cultural group membership, as well as exercising a powerful effect in its own right. Frierson (1965) compared the characteristics of children from lower socioeconomic backgrounds to those from more favorable circumstances on aspects of superego development, activity preferences, and creative thinking. He found that gifted disadvantaged children tended to demonstrate less superego development, preferred participation in games and competitive sports over reading, and were inferior in terms of creative thinking than their more advantaged counterparts. While these findings are obviously generalizations, the trends indicated would clearly be more of a hindrance than a help to the socioeconomically disadvantaged when it comes to academic performance.

First Movement §1(c): The role of home values

Lest the weight of the above observations lend a sense of inevitability to the effect of socioeconomic disadvantage, the influence of the home has long been noted in modulating such effects. Some studies have shown the home influence to be negative. Entwistle and Hayduk (1978) pointed out the differences in expectations of academic success between the children from working class backgrounds compared to those from middle class backgrounds, with working class children setting unrealistically high expectations, thereby setting themselves up to experience failure. The working class parents entertained more realistic expectations than their children, but were still less accurate in their predictions than were the middle class parents of their children. McIntosh and Greenlaw (1986) found that lower socioeconomic homes communicated different values in relation to academic achievement than did middle and upper class

homes. Lower socioeconomic homes were likely to devalue education per se, to value holding a job more than pursuing a career, to hold that postsecondary education was not necessary, and to focus on the present rather than planning for the future.

In contrast, other studies have shown the home to be a very positive influence. Hanson and Ginsburg (1986) found that high expectations correlated with high achievement patterns in socioeconomically disadvantaged children. Values exerted twice as strong an effect as did socioeconomic status in determining school success. Factors such as high parental expectations, peers who value education, personally high educational expectations, and being in control of one's own future were all positively associated with increase in achievement over time. VanTassel-Baska (1989) highlighted the role of family and extended family members of successful gifted disadvantaged children in stressing the value of education and the work ethic, and monitoring the child's education. Borland, Schnur and Wright (2000) reported the results of a follow-up to an earlier study (Borland & Wright, 1994) in which a number of students were identified from a socioeconomically disadvantaged area of Harlem, New York. While the sample size was quite small (5 students), the ethnographic methodology utilized was rigorous in stressing the role that the home played in the success of the Project Synergy intervention. It was summed up in eight assertions concerning the "successful" families as follows:

1. The parents do not believe in a totally intractable caste system.
2. Parents believe that academic success can lead to upward mobility and socialize their children accordingly.
3. The parents create a home environment in which the prevailing norms resemble middle-class norms, lessening the pressure on the children for "assimilation without accommodation."

4. The parents are unwilling to attribute all disappointments to racism or are willing to ignore some instances of racism for their children's sake.
5. The parents recognize and encourage their children's giftedness.
6. There are positive role models for the children in the home, including parents.
7. The parents are willing to take risks.
8. There is no overt family pathology. The families, although headed by a single mother in four of five cases, are stable and provide love and support for the children. (Borland, Schnur, & Wright, 2000, p. 26).

First Movement §1(d): Early intervention desirable

All of this leads one to suggest that early intervention is desirable. VanTassel-Baska (1991), drawing from studies by Ramey, Yates, and Short (1984), and Seitz, Rosenbaum, and Apfel (1985), declared that early intervention "has been influential in reducing later academic problems for disadvantaged students" (p. 84). In reviewing the effect of Head Start programs, Lazar (1981) found that participants were significantly more likely to finish high school, stay out of special education, and complete high school on time than their socioeconomic peers who were not involved in Head Start. In a study of children in preschool programs in the 60's and 70's, Royce, Lazar, and Darlington (1983) had found that the earlier the educational intervention, the more likely it was to be effective, and that small adult-child ratios, parental participation, and working with the family situation rather than just with the child all contributed to beneficial outcomes. In a more recent study, Marcon (1999) found similar effects with three cohorts of preschool children, most from low-income, single-parent, hard-to-engage families. Marcon (1990) found that increased parent involvement had a positive impact on the preschoolers' early development and mastery of basic skills.

Vinovskis (1999) addressed the broad question of the effectiveness of federal compensatory education programs, especially the young children emphasis of Title 1 and Head Start. His verdict was guarded, focusing on the fact that “major evaluation studies have repeatedly found at best that only modest gains result from these programs” (p. 197). However, Vinovskis (1999) went on to cite a far more optimistic evaluation by Barnett (1998), who criticized a number of what he regarded as deficiencies in a number of key studies in commenting that:

for economically disadvantaged children ECE (early childhood education) substantially improves cognitive development during early childhood and produces long-term increases in achievement (learning) and school success. The evidence of long-term effects is provided by thirty-eight studies and generalizes across a wide range of programs and communities. Although many studies fail to find persistent achievement effects this is plausibly explained by flaws in study design and follow-up procedures. (Barnett, 1998, p. 38)

Vinovskis (1999) instanced the Perry Preschool Program as an outstanding example of a successful program, and noted that “the high quality of care provided to the youngsters does not resemble that of most Head Start programs” (p. 196). Far from helping the children most in need, Vinovskis asserted that “unfortunately the children who were the most disadvantaged (and a particular focus of Title 1 funds) were not helped much at all” (p. 190). This clearly pointed to the need to consider the means of identification of young high-ability learners.

First Movement §2(a): Traditional methods for identifying high-ability learners
Gottfried, Gottfried, Bathurst and Guerin (1994) predated Kirschenbaum (1998), Gagne (1997), Sternberg (1998), among many others, in asserting that “most school

districts still rely on traditional assessment methods” (p. 140) to identify high-ability learners. There is a vast array of instruments and tests that can be used singly and in combination to assess the learning ability of children. Ability tests (both group and individual) and achievement tests (both group and individual) are commonly used as gatekeepers to programs for high-ability learners. While by far the greatest amount of testing in 1993 was related to achievement (Figure 4), a range of other instruments including inventories (like behavioral checklists), creativity tests, syntheses of grades, portfolios, and interviews were also used.

A result on a particular test is an indicator only. “Although they are helpful tools, *test results should never be used as the sole determiner for any educational decision*” (Harcourt Educational Measurement Inc., <http://www.bjup.com/testing/successfaq.html>). It is the prerogative of the local school district to set the parameters for identification, in accord with relevant state directives. The Virginia Plan for the Gifted (1996), for example, is one state’s attempt to offer some guidance as to what might be appropriate at

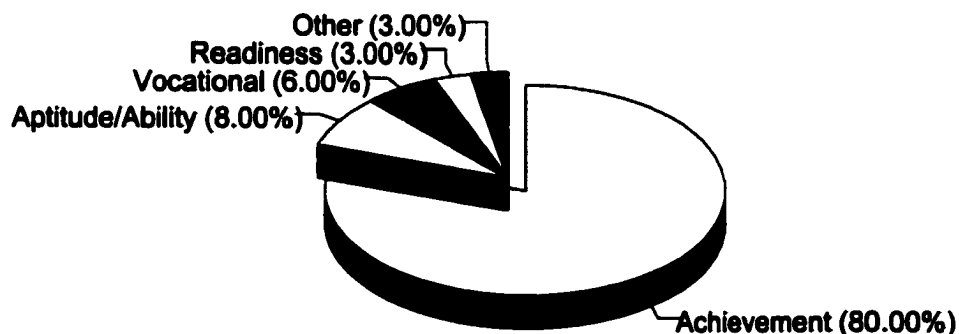


Figure 4 Types of standardized tests in American schools. (USGAO, 1993, p.21)

various levels. It lists thirty-eight “frequently used assessment instruments,” of which seventeen are designated as “effective in identifying potential in special populations.” These special populations were defined as encompassing those from low income and culturally diverse backgrounds. Of this seventeen, nine were described as being both verbal and nonverbal in orientation, and one—Scales for Rating Behavioral Characteristics of Superior Students (Renzulli, Smith, White, Callahan, & Hartman, 1976)—was listed as verbally oriented. The other seven in this set were described as being nonverbally oriented.

**First Movement §2(b): Academic disagreement concerning the meaning of
outcomes**

Quite apart from the potential for conflicting interpretations of the meaning of one individual’s scores on different tests across the range of available tests, it is informative to consider how even different versions of the same test can yield divergent scores. Such potential divergence casts doubt on the outcome of testing for the purpose of identification of high-ability learners in general, and, more specifically, for the issue of the identification of high-ability learners from among the socioeconomically disadvantaged. A brief discussion of the Weschler Intelligence Scale for Children (WISC III), one of the more widely known and respected tests of general ability, will serve to illustrate this point.

The WISC-III is designated as appropriate for ages 6 years to 16 years 11 months. Fishkin, Kampsnider and Silverman (1997) reviewed and summarized the outcomes of seven published studies on the WISC-III. Fishkin et al. concluded that “to be gifted on the WISC-III, children must be adept and quick. Those children who were reflective (were) unable to earn the bonus points to score in the gifted ranges” (Fishkin, Kampsnider &

Silverman, p. 2). Fishkin et al. (1997) suggested that the emphasis on speed, which Kaufman (1992) had earlier described as “excessive” and “foolish” (p. 157), may well have accounted for the lower FSIQ (full scale IQ) scores of academically gifted youth measured with the WISC-III as opposed to the WISC-R, the Binet L-M, or the Binet-IV. In summation, Fishkin et al. (1997) commented that their results supported the growing body of evidence against the use of the WISC-III FSIQ as the primary criterion to identify gifted levels of ability.

When attention is directed to the WISC-R (the revised version which preceded the third edition) rather than the WISC-III, the same type of discrepancy among the subtest scores is noted. Intellectually gifted students often show lower performance than verbal IQ scores, with an attendant depression of the FSIQ (Silver & Clampitt, 1990). According to Silver and Clampitt, discrepancies as large as 21 points were not at all rare in the academically gifted population, occurring in at least one-fifth of the children whose verbal or performance IQ was greater than 130. A number of other researchers (Brown & Yakimowski, 1987; Hollinger & Kosek, 1986; Patchett & Stansfield, 1992; Wilkinson, 1993) have also commented that gifted children also often showed considerable deviation among their subtest scores on the WISC-R. Hence, use of the WISC-R does not in itself yield a more consistent identification protocol for high-ability learners.

Silverman (1997) has suggested that verbal competency measures such as the Verbal Comprehension Index (VCI) of the WISC-III should be used in preference to the WISC-III FSIQ as an identification instrument. Fishkin (1997) ventured that the VCI identified a more coherent group of children, and that the abilities clustered by the VCI was conceptually more coherent by virtue of the factor analytic basis of the score. Fishkin (1997) concluded that “evidence of superior abilities on those abilities comprising the

VCI are consistent with characteristics that have traditionally been clearly recognizable as intellectually gifted abilities” (p. 5). Fishkin does not address the issue of what those recognizable “intellectually gifted abilities” might be, but one of the issues arising out of this discussion is that if the VCI is used to identify children from within the underserved populations, then the fairness of the identification procedure is questionable.

First Movement §2(c): Appropriate tests for high-ability learners

The above discussion of the WISC family of tests illustrates that there are a number of cogent issues concerned with the outcome of testing, and that these very issues validate the call for another dimension besides general aptitude in an identification protocol. Indeed every test has characteristics which make it more suited to one particular testing niche than to others. School districts in South Carolina commonly use the Otis-Lennon School Ability Test (OLSAT), the Test of Cognitive Skills, the Standard Progressive Matrices, and the Terra Nova test as tests of aptitude (with the Otis-Lennon by far the most common), and the Metropolitan Achievement Test (MAT-7), the Comprehensive Tests of Basic Skills, and the Terra Nova test as tests of achievement (with the Metropolitan Achievement Test by far the most common).

Because the OLSAT and the MAT-7 were by far the most commonly used tests of ability and achievement respectively in South Carolina it seemed appropriate to briefly overview each before moving to suggest that they were ineffective in the context of this study.

The Otis-Lennon School Ability Test (OLSAT)

Anastasi (1992) set the sixth edition of the OLSAT (the edition relevant to this study) in context as the latest of a series of tests that “virtually spans the history of group testing, from the pioneering innovations of Otis in 1918” (p. 633). Swerdlik (1992)

quoted the OLSAT Technical Manual when he declared that the major purpose of the test was to “assess examinees’ ability to cope with school learning tasks, to suggest their possible placement for school learning functions, and to evaluate their achievement in relation to the talents they bring to school learning situations” (p. 636).

Starting with the sixth edition, the OLSAT total score was called an SAI for “school ability index.” These are normalized standard scores ($M = 100$, $SD = 16$) within each 3-month age group from 4 years 6 months to 18 years 2 months. The total SAI has two component parts of the verbal score (verbal comprehension and verbal reasoning), and two of three components to the nonverbal score, depending on the test level, chosen from pictorial reasoning, figural reasoning and, quantitative reasoning. The nonverbal items, according to Anastasi (1992) involved “essentially comprehension and reasoning with nonverbal content” (p. 634).

The OLSAT was a very carefully normed test—normed on a sample as nearly representative of the American school-age population as could be obtained, based on the 1980 census data. Special care was taken to include non-public school children, and children with “various physical and psychological handicaps” (Anastasi, 1992, p. 635). Swerdlik (1992) obliquely highlighted the difficulty in a number of districts in which OLSAT is used, when he issued “specific cautions against any possible misuses of the test such as educational placement based solely on the test scores” (p. 636). In other words, the difficulties start to arise when this well-credentialed test is used for a purpose for which it was not intended.

Metropolitan Achievement Test (MAT-7)

“Historically, standardized achievement tests have performed well: they efficiently provide accurate information about students’ skills in areas such as reading

comprehension, mathematical computation, locating and using resource materials, and placing correct punctuation in a sentence” (Cisek, 1998, p. 2). As Cisek goes on to point out, “the picture becomes more complex when tests are used to gauge the learning of groups of students” (p. 5).

The Metropolitan Achievement Test is an established achievement test moving into its eighth edition in Fall 2000—an edition which is being touted as a test which implies “specific action strategies for teachers and parents” (Harcourt Educational Measurement Inc., <http://www.hbem.com/trophy/achvtest/mat8.htm>). The MAT-7 (the edition relevant to this study) covers four content areas: reading, mathematics, language, and other. Reading is broken down into three subdivisions at the early elementary level (vocabulary, comprehension, and word recognition), with word recognition being dropped for the late elementary version. Mathematics consists of two subdivisions: concepts and problem solving, and procedures. The one Language division for the early elementary is subdivided into prewriting, composing, and editing at the late elementary stage. The “other” division subsumes science and social studies at the early elementary level, while thinking skills and research skills are added at the late elementary stage. In all, there are 14 different levels of the test, ranging from youngest kindergarten to oldest high school.

The development of the MAT-7 “faithfully followed standard procedures of test development starting with a review and analysis of recent editions of major text-book series in every subject area covered by test batteries” (Finley, 1995, p. 603). While textbooks might well provide a convenient place to start, there is good reason to doubt the wisdom of using an instrument arising from such a source with high-ability learners. It has long been known that textbooks not infrequently fail to stimulate the high-ability

learner. For example, Renzulli, Smith and Reis (1982) reported a study by Educational Products Information Exchange (EPIE) which found that over half of the fourth graders in some school districts were able to achieve a score of 80% or higher on a test of the content of their math texts before they opened their texts in the fall. Similar results were found in science and social studies texts (Gallagher and Gallagher, 1994). Venezky (1992), in an extremely comprehensive survey of textbooks in the light of their role in society, pointed out the non-educationally based imperatives that govern the content of textbooks and the way in which material is treated. He presents cogent arguments to support a belief that better teaching happens when textbooks are only minimally in evidence. Aside from these two reasons for expressing reservations about a testing approach arising from a textbook survey, two other important problems were raised by Hambleton (1995), who questioned how effective a textbook review would be in reflecting the growing curriculum diversity, and the effectiveness of the multi-choice answer format is in assessing higher-order thinking skills: "The validity of the multiple-choice item format for assessing many important school outcomes has been seriously challenged by many educators" (p. 607). What is being pointed out here is not that the MAT-7 has no role in testing, but that it certainly has drawbacks in terms of assessing the very characteristics one would expect to find in the high-ability learner.

First Movement §2(d): "Traditional" procedures are ineffective

Lidz (1987) declared "psychologists have long expressed dissatisfaction with traditional models of assessment and have called for change" (p. 3). Passow and Frasier (1996) did not tread lightly in declaring that "the most widely accepted explanation for the low participation of disadvantaged students in programs for the gifted is the ineffectiveness and inappropriateness of the identification and selection procedures that

traditionally have been and continue to be used” (p.198). As was noted in National Excellence (OERI, 1993), states that use IQ score cutoffs to identify gifted and talented students are more likely to have large disparities among racial and ethnic groups.

The concept of what constitutes effectiveness needs to be addressed, because it is clear that the procedures historically utilized have indeed identified some high-ability learners. For example, when Terman placed emphasis on selecting the youngest and brightest students in the class for inclusion in his study, he certainly did select a group of high-ability learners. But even if this somewhat arbitrary selection process was sufficient for his research purposes, such a selection procedure is unsuitable in an educational setting where one has to have a defensible basis for making decisions concerning access to educational services. It is not sufficient to claim one is educating “some” high-ability learners and ignoring the ones who don’t surface. In contrast, if a process can be devised and can be shown to identify high-ability learners precisely as a result of a performance demonstrating their enhanced ability to learn, then the “defensibility”—the fairness—of the process is inherently demonstrable.

**First Movement §2(e): Regulations covering gifted identification in South
Carolina**

Regulation 43-220, Gifted and Talented, of the State of South Carolina

Department of Education regulations was amended on May 12, 1999 and published in the State Register on May 28, 1999. The new regulations make provision for the continuing cohort of identified students, and for children transferring in to one district from another South Carolina district where they were already identified. The definition of the population to be served was set as students who “meet the 96th national age percentile composite score or higher (placement grades 3-12) or the 98th national age percentile

composite score or higher (placement grades 1-2) on an individual or group aptitude test ” (Lee & Lord, 1999, p. 5). However, the new regulations went on to provide that a student could be identified if he/she met “the criteria in two out of three dimensions that follow” (Lee & Lord, 1999, p. 5).

Dimension A was designated as “reasoning abilities” and set the bar at “high aptitude (90th national age percentile or above) in one or more of these areas: verbal/linguistic, quantitative/mathematical, non-verbal, and/or a composite of the three” on either an individual or group aptitude test (Lee & Lord, 1999, p. 7). Dimension B was designated “high achievement in reading and/or mathematical areas” and encompassed “high achievement (94th national percentile and above or advanced status) in reading and/or mathematical areas as measured by nationally normed or South Carolina statewide assessment instruments” (Lee & Lord, 1999, p. 7). Dimension C was entitled “intellectual/academic performance” and nominated students who “ demonstrate a high degree of interest in and commitment to academic and/or intellectual pursuits or demonstrate intellectual characteristics such as curiosity/inquiry, reflection, persistence/tenacity in the face of challenge and creative productive thinking. Characteristics for this dimension are demonstrated through: ...(b) assessments of performance tasks for placement in Grades 1-6...The performance standard is four points on a five point scale” (Lee & Lord, 1999, p. 7-8).

Given the issues raised in the previous sections discussing the OLSAT and MAT-7 tests, it is clear that South Carolina has much to gain by adding Dimension C-- performance-based identification--as the third strand of identification in addition to the data from Dimension A and Dimension B. Clearly there is no reason to suppose that socioeconomically disadvantaged children could not be in the very highest aptitude and

achievement bands, although evidence presented in the following section points to such an outcome as being unusual. The major concern in this study, however, is to make it distinctly possible for socioeconomically disadvantaged children who qualify on only one of Dimension A or Dimension B to be able to avail of the Dimension C performance-based identification protocol—which bodes well to minimize their disadvantage—to gain qualification.

**First Movement §3(a): Alternative ways of identifying children from
socioeconomically disadvantaged backgrounds**

There is evidence that socioeconomic disadvantage depresses the performance of socioeconomically disadvantaged students on standardized tests. This evidence provides a strong argument for alternative assessment protocols. For example, VanTassel-Baska and Willis (1987) found socioeconomic disadvantage, as defined by low income, was a factor in accounting for the lower scoring by such students on all sections of the Scholastic Aptitude Test (SAT), regardless of ethnicity. The Alamprese, Erlanger, and Brigham (1988) study found that socioeconomically disadvantaged students comprised 20% of the student population, but made up only 4% of those who performed at the highest levels on standardized tests.

Borland, Schnur and Wright (2000) declared that “even without additional funds, schools can do more to identify giftedness among economically disadvantaged students than they are now” (p. 27). They saw schools as needing to be involved in nontraditional identification procedures and cited portfolio assessment (Wright and Borland, 1993), focusing on best performance (Roedell, Jackson, & Robinson, 1980), and dynamic assessment (Lidz, 1987) as providing alternative models. Frasier (1993) discussed the advantages and difficulties associated with a checklist approach.

Portfolio identification

Wright and Borland (1993) ventured that student portfolios had become a major topic in education, but noted regretfully that there were few examples of what they called serious examples of the use of portfolios by educators of gifted students. They pointed out that portfolios were being used far more for assessment than for identification purposes—despite the fact that portfolios held great promise for identification among “two overlapping populations, young children and economically disadvantaged students” (Wright & Borland, 1993, p. 205). The main advantages they saw in portfolio identification included the “ongoing, ecological, and curriculum-focused” nature of such a regime, the eschewing of “one-time psychometric assessments,” and their perception that in ways not specifically enumerated, portfolios provided “a way to overcome the problems encountered” in identification among potentially gifted children who are economically disadvantaged (p. 205). Paulson, Paulson and Meyer (1991) underlined the potential for portfolio identification when they declared that “portfolios have the potential to reveal a lot about their creators. They can become a window into the students’ heads, a means for both staff and students to understand the educational process at the level of the individual learner” (p. 61).

Wright and Borland (1993) followed their own advice and used the portfolio methodology to identify the children involved in Project Synergy (Wright & Borland, 1993; Borland & Wright, 1994), explicitly delineating the contents of what they called the Early Childhood Developmental Portfolio as “a systematic compilation of selected examples of a child’s work and records of observations of a child’s behavior that document the child’s status and growth in one or more developmental domains” (Wright & Borland, 1993, p. 206). Wright and Borland (1993) went on to suggest that their Early

Childhood Developmental Portfolios contain three kinds of work samples, “compiled by collecting photographs, audio tapes, video tapes, and the children’s work itself” (p. 206).

At a more general level, Johnson (1996) described three variations on the portfolio assessment theme: best-works, selection, and process. As Johnson described it, best-works portfolios “show off the exemplary work of the person submitting it and the choices as to what is submitted are made by the person presenting the portfolio” (p. 30). For the selection portfolio, the decisions concerning what material is to be included are made by the person who is compiling the portfolio in conjunction with the person to whom the portfolio is ultimately to be submitted. Finally, the process portfolio is intended to show “a span of work from an early stage to a finished product” (p. 30). One of the key points that Johnson (1996) made is that the purpose for compiling the portfolio must stay at the forefront of the whole enterprise—an admonition elegantly illustrated by Wright and Borland (1993).

Ingels and Quinn (1996) wrote enthusiastically about the potential for portfolio assessment to redress imbalances in identification, though they added a caveat relating to the labor-intensiveness of the process. They commented that portfolio assessment empowers one to be “very inclusive indeed, though being so may prove expensive” (p. 43).

Best performance

Roedell, Jackson and Robinson (1980) urged those who are looking to identify children who are high-ability learners to look at the best work that the student has produced, and not to be swayed by an inclination to, even inadvertently, average out the totality of the observed performance. The rationale for looking at the best effort, particularly for disadvantaged children, arises from an acknowledgment that students

from such backgrounds need to be given the benefit of the doubt in view of the way in which the effects of a disadvantaged home background may interfere with their ability to perform consistently at a high level. For example, Passow (1982) nominated a number of debilitating factors for these children, including experiential deprivations (especially in early childhood), limited language development, and socioeconomic or racial isolation.

Direct observation

Direct observation has been recommended by Chittenden (1991) as “potentially the richest source of information” (p. 25). Observation, in this context, consists of the sort of information that teachers note in everyday work with children; that is, cues in children’s language and behavior that signal their interests, their thinking, their relationships. This category includes too the children’s own observations and ideas about their works. (Chittenden, 1991, p. 25)

Wright and Borland (1993) included direct observation as one of the two basic strategies constituting their Early Childhood Developmental Portfolio. They commented further that “for richness of detail, sensitivity to change over time, and potential validity, direct observation of student performance and behavior is unparalleled as a means of assessment” (p. 206).

Dynamic assessment

In a masterful historical overview of the concept of dynamic assessment, Lidz (1987) nominated two words as being of primary importance to its definition and conceptualization: activity and modifiability. Lidz commented: “The examiner and learner are both active; the examiner is an active intervener who monitors and modifies the interaction with the learner in order to induce successful learning” (p. 3). Lidz continues:

Dynamic assessment, then, is an interaction between an examiner-as-intervener and a learner-as-active-participant, which seeks to estimate the degree of modifiability of the learner and the means by which positive changes in cognitive functioning can be induced and maintained. (p. 4)

The strength of dynamic assessment as a methodology in the identification of high-ability learners in the context of socioeconomically disadvantaged children lies in the fact that there are minimal presumptions made with respect to any prior learning. The tasks used to detect high-ability learning can be made highly culture-free, and even reading ability becomes less significant, a strength shared with the Raven (1987) series of tests. In short, the advantages of the dynamic assessment methodology for identifying high-ability learners are the same as for identifying struggling learners—only the tasks need to be different. This discussion of dynamic assessment will be continued in the context of Feuerstein’s implementation of the technique in the second movement.

First Movement §3(b): Other factors to be considered

VanTassel-Baska (1991) delineated four key issues in relation to identifying high-ability learners from disadvantaged populations. In addition to using nontraditional measures to identify disadvantaged students, as discussed above, the other three are “recognition of cultural attributes and factors in deciding on identification procedures, ...a focus on strengths in nonacademic areas, particularly in creativity and psychomotor domains, ... creation of programs that address noncognitive skills and that enhance motivation” (p. 80).

Borland, Schnur and Wright (2000), after commenting on the need for alternative assessment procedures, opined that schools will also require “an understanding that giftedness manifests itself in different ways in different settings, and that, in order to

understand these manifestations, one must understand the setting” (p. 28). These opinions provide an ideal segue into the consideration of the ideas of Kenneth Burke, as modulated by Wertsch (1998), in relation to what Burke termed “dramatism.” However, to do so would be to conflate the two overarching themes of this chapter before the second theme has time to develop, and before the sweep of the first theme has been reinforced. Consequently discussing the link to dramatism will be delayed until the coda of the second theme. It is fitting to conclude this movement by briefly recapitulating, and then underlining the main issues in the distinction between development and learning which was interwoven throughout the discussion to date. This will naturally lead the discussion back to the concept of dynamic assessment at the start of the second movement.

First Movement §4(a) The relationship between development and learning

The concept of dynamic assessment is closely linked with the work of Vygotsky (1978) who described one’s facility to learn when confronted with stimuli as being commensurate with one’s zone of proximal development. The zone of proximal development was the key to Vygotsky’s epistemology. Each of three alternative positions to Vygotsky’s concerning the relationship between development and learning contribute to the background against which the richness of the zone of proximal development emerges.

Vygotsky grouped the conceptions of this relationship current in his day into three major groups. The first of these three theoretical groups subscribed to the basic tenet that the processes of child development were independent of learning. Chief among the proponents of this position was Jean Piaget, although, according to Vygotsky (1978), many of the classics of psychological literature including the works of Binet are in a similar vein. In Glassman’s (1994) opinion that “there is little doubt that Piaget’s work

informs Vygotsky's theoretical perspective...and there is little doubt that Piaget finds an affinity between his theory and Vygotsky's work" (Glassman, 1994, p. 186). From Vygotsky's (1978) point of view, the proponents of the Piagetian perspective held that learning utilized the achievements of development rather than providing impetus for modifying its course. Processes of deduction, understanding, logical thought, and similar thinking processes were seen by Piaget as essentially occurring by themselves. Formal schooling, for example, was seen as having little effect on the developmental stage or the sequence of the stages. For example, the point of asking a child "why doesn't the sun fall?" is to present the child with a question the answer to which he or she has no ready access. Neither does the child possess the general capabilities for generating an answer. As a consequence of this dearth of knowledge the child is compelled to answer without reference to prior learning, and so the questioner is able to gain a clearer picture of the child's thinking tendencies. Vygotsky spoke of proponents of the Piagetian point of view as "especially (fearing) premature instruction, the teaching of a subject before the child (is) ready for it" (Vygotsky, 1978, p. 80). Vygotsky summarized this theoretical approach as considering that "learning forms a superstructure over development, leaving the latter essentially unaltered" (Vygotsky, 1978, p. 80).

The second major theoretical position regarding development and learning, as nominated by Vygotsky (1978), was that maintained by the theoreticians who held that learning is synonymous with development, that the two concepts were inseparably linked. Vygotsky nominated the work of William James as being typical of this position. James expressed his position thus: "Education, in short, cannot be better described than by calling it the organization of acquired habits of conduct and tendencies to behavior" (James, 1958, p. 36-37). Education organizes that which maturation provides. What one

learns is to marshal the inclinations and motivation provided by the sheer weight of experience. Vygotsky saw this second position as the essence of a group of theories of diverse origins, and he referred to these theories as being what he called “reflex” theories. In an evaluation of reflex theories which was echoed a number of times, Vygotsky designated common ground between this second group and the Piagetian position that “development is ... the elaboration and substitution of innate responses” (Vygotsky, 1978, p.80).

To summarize Vygotsky’s position so far, he proposed that the first view held that developmental cycles precede learning cycles, and that maturation must therefore precede learning, with instruction appropriately lagging behind mental growth. For the second group, both “learning and development occur at all points in the same way that two identical geometrical figures coincide when superimposed” (Vygotsky, 1978, p. 81).

The third position on the issue of the relationship between development and learning was characterized by Vygotsky (1978) as being a simple combination of the two preceding positions. He nominated the work of Koffka as being representative of this third group. Vygotsky understood Koffka’s position as being that development was the outcome of a dynamic equilibrium between maturation (which depended on the development of the nervous system) and learning (which was also a developmental process). Vygotsky took heart from Koffka’s synthesis in that its success demonstrated a degree of compatibility between the first two approaches. Vygotsky characterized the interplay of learning and development as a step toward an increased level of understanding, and pointed out that the problem of transfer of learning was brought into sharper focus by Koffka’s synthesis. In particular, the dubious validity of the assumption

that “mental capabilities function independently of the material with which they operate, and that the development of one ability entails the development of others” (Vygotsky, 1978, p. 82) was strongly highlighted—the very point which Thorndike (1914) had pointed out much earlier.

Having proposed his interpretation of these three positions, Vygotsky (1978) rejected them all, and instead framed his own position by referring immediately to formal schooling as the touchstone. It is not that his position only applies to the child who experiences formal schooling, but it is in this milieu and among the dimensions of school learning that he introduces his theory. His position, Vygotsky declares, is close to, but distinct from Koffka’s. Whereas Koffka “and others assume that the difference between preschool and school learning consists of non-systematic learning in one case and systematic learning in the other” (Vygotsky, 1978, p. 84-85), Vygotsky avers that “learning and development are interrelated from the child’s very first day of life” (p. 84). School-based learning was not just more systematic learning than that which the child’s environment had provided in the preschool years. It “introduces something fundamentally new into the child’s development” (p.85). The fundamentally new mechanism which Vygotsky proposed as the driving force of learning and the foundation of his theoretical position was mediated by the interaction between the child, adult teachers and more knowledgeable peers, and was called the zone of proximal development.

In essence, Vygotsky’s (1978) theory holds that it is neither the child’s actual development as measured by the child’s completed developmental cycles, nor the child’s potential development, as measured by what the child can do with the assistance of an adult teacher or knowledgeable peer, that is significant, but the difference between the two. Vygotsky was somewhat dismissive of the efforts of those who placed their faith in

conventional ability measurement which he conceived of as being based on the assumption that actual development is the best measure of mental ability. “Over a decade even the profoundest thinkers never questioned the assumption; they never entertained the notion that what children can do with the assistance of others might be in some sense even more indicative of their mental ability than what they can do alone” (Vygotsky, 1978, p. 85).

First Movement §4(b) The dialectical nature of growth

A number of authors have compared and contrasted the theories of Piaget and Vygotsky. Tudge and Winterhoff (1993) summarized the typical positions ascribed to each as follows:

Vygotsky believed that development, a social process from birth onward is assisted by others (adults or peers) more competent in the skills and technologies available to the culture, and that development is fostered by collaboration within the child’s zone of proximal development. Piaget believed that children are like scientists, working alone on the physical, logical, and mathematical material of their world to make sense of reality. To the extent that they can benefit from interaction, it is with peers rather than adults, the dominant mechanism driving development being “cognitive conflict.” (p. 62)

Glassman (1994) went further to assert that “Piaget’s equilibration theory and Vygotsky’s socio-historical framework are actually closer than is usually recognized” (p. 186). Glassman depicted the essence of Vygotsky’s “frustration and despair” (p. 187) with Piaget as stemming from the latter’s falling victim to “the cruel fate of idealism” (p. 187) in abandoning his reliance on the real world, and reverting to a Freudian model in relation to the genesis of language. This is a fine point, and one which Glassman was

careful to nuance, but on which he maintained a neutral stance. The major problem for Vygotsky was that by reverting to idealism, Piaget avoided making a judgement about whether there is a primary cause in development. "Vygotsky claims that Piaget sees development as an unending stream in which A and B are dependent on each other, but there is no way to posit the initial cause" (Glassman, 1994, p. 188). In contrast, Vygotsky strongly posited "social interaction as the necessary and primary cause of the ontological development of knowledge in the individual" (Glassman, 1994, p. 188).

In contrast, little attention has been paid to Piaget's interest in children's involvement in their social world, nor to Vygotsky's discussion of the impact of maturational factors, or his statements concerning the role of imitation, which, he said, was "the source of instruction's influence on development....Instruction is possible only where there is potential for imitation" (Vygotsky, 1987, pp. 210-211).

Both Vygotsky and Piaget believed in the dialectical nature of development (Tudge & Winterhoff, 1993). Both believed that there was an inherent tension developed in the individual in the process of learning. For Vygotsky it was the tension between the actual developmental level and the potential developmental level. This tension preserves the more or less permanent zone of proximal development. van Geert (1998) pointed out that one of the mechanisms not specified in Vygotsky's theory was one which could account for the re-expansion of the actual and potential levels after a learning episode had narrowed the zone of proximal development. For Piaget the tension lay in the creation of disequilibrium at the point where the accommodative response could stretch no further. This led to the breaking down of formerly adequate cognitive schema, and the emergence of new structures by means of assimilating the new concepts into the old to generate a new cognitive schema.

Both Vygotsky and Piaget acknowledged their intellectual debt to J.M. Baldwin. Vygotsky approved of Baldwin's declaration that "the task of genetic psychology was to 'specify those forms of social interaction which enable individuals to develop'" (quoted in Tudge & Winterhoff, 1993, p. 64). Piaget asserted that "as J.M. Baldwin saw quite clearly, the formation of the self is connected to early interpersonal relationships and especially imitation" (quoted in Tudge & Winterhoff, 1993, p. 64). In Piaget's case there is an even greater debt to Baldwin, as "the very concepts of assimilation, accommodation, and equilibration are all to be found in Baldwin's writings" (Tudge & Winterhoff, p.64).

In summary, this first movement began by establishing a basic foundation of concern about the education of the socioeconomically disadvantaged. It moved on to develop a perspective on current identification practices and pointed out that alternatives that address a number of concerns do exist. The latter part of this movement set the earlier discussion in a theoretical matrix that has its roots in the thoughts of some of the giants of the educational field. The view that development, or maturation, is essential before appropriate material can be learned (a loosely Piagetian perspective) is antithetical to the approach taken in this study, which was built on a Vygotskian learning paradigm. It was believed to be important to clearly establish the theoretical position out of which this identification endeavor grew.

Returning from this main theme of the first movement of this chapter, the second movement begins by a quick reprise followed by the development of the concept of dynamic assessment.

Second Movement §5(a): Feuerstein and dynamic assessment

Lidz (1987) traced the earliest approaches to dynamic assessment back to the 1920s, and highlighted the changes and development of the concept by decades up to her day, highlighting the work of Feuerstein, Budoff, Campione and Brown and Stott and their differing applications of the concept in the 1970s. “It was also in the 1970's that Vygotsky’s proposed ‘zone of proximal development’ was realized in assessment procedures developed by Campione and Brown” (Lidz, 1987, p. 16).

Minick (1987) agreed with Campione, Brown, Ferrra and Bryant (1984) in the assertion that

attempts to develop dynamic assessment procedures have consistently been motivated by the conviction that static approaches to the assessment of learning ability or learning potential have failed to provide the kinds of information that educators need in order to facilitate the psychological development and the educational advancement of these children. (Minick, 1987, p. 116)

Minick (1987) went on to discern two distinct traditions within the dynamic assessment movement. One was referred to as the quantitative tradition based on a test-train-retest format. In this tradition, associated with Brown, Campione and Budoff, after establishing a baseline, “the examiner provides a controlled protocol of assistance and instruction while the child is working on comparable tasks. Finally, the child is observed while working alone to assess the amount of benefit from (the) instruction” (p. 117).

The second tradition, characterized by Minick (1987) as qualitative, queries the need to establish a baseline in the first instance. The child who fails on the test has a negative perception of the task which makes it more difficult for the examiner to engage him or her in subsequent work on similar tasks. This approach was typically associated

with Feuerstein (1979) who also insists on maintaining flexibility in the examiner's interaction with the child. The lack of a baseline measure and the non-standardized interaction protocol combine to lessen the reliability of any quantitative measures of Feuerstein's interventions—a shortcoming which Feuerstein is happy to tolerate because of the more useful information (in terms of the child's psychological processes and information on the type of help most likely to be of benefit) obtained through the more qualitative approach.

Both the quantitative and qualitative traditions trace their theoretical foundations from the work of Vygotsky (1978), but it was Minick's (1987) judgement that Vygotsky's concept of the ZPD, and the system of theory and research of which it is a part, have more direct implications for the kinds of assessment problems that have been addressed in the work of Feuerstein and his colleagues than they have for the task of producing quantitative measures of a child's learning efficiency or learning potential. (p. 119)

The problem of identification of children able to learn from among a diverse population confronted Feuerstein with some immediacy as the fledgling state of Israel opened its doors to Jewish immigrants from a multiplicity of European countries in the aftermath of the holocaust. Many of the children had come from ghetto-like situations in which they had little opportunity for education in any formal sense. Hence, when assessed in any traditional way, many appeared to be academically retarded in the sense that they appeared to be unable to benefit from the usual teaching environment. Feuerstein sought to discriminate between having knowledge deficits as opposed to having intellectual deficits; between being unable to show evidence of learning as opposed to being unable to learn. His task was complicated because many of the children

did not speak the languages in which standardized tests were written.

One of the ways in which this problem was approached by Feuerstein and his associates resulted in the production of the Learning Potential Assessment Device (LPAD; Feuerstein, Rand, & Hoffman, 1979), which is “still the only comprehensive test of learning ability that uses the dynamic assessment method exclusively” (Kirschenbaum, 1998, p.141). According to Sternberg (1993), in the LPAD, “an examiner gives children rather difficult tasks to solve. Initially, he or she looks at how the children solve the tasks without any intervention on the part of the examiner. Then, children receive carefully graded, sequential hints, and the examiner observes the children’s ability to profit from these hints. In this way it becomes possible to observe the children’s zone of proximal development” (p. 202).

Second Movement §5(b): Another look at the zone of proximal development

Vygotsky is credited by Sternberg (1993) as largely motivating the sociological approach to intellectual potential. In Sternberg’s view, the most important two contributions by Vygotsky to the theory of intelligence were his theory of internalization and his conceptualization of the zone of proximal development. To briefly reiterate some of the discussion above, in his theory of internalization Vygotsky started from a premise which is exactly the opposite of that taken by Piaget. Both believed that intelligence developed by means of interaction with the environment, but Vygotsky believed that intelligence begins in the social environment and directs itself inward by means of the process of internalization. Piaget, on the other hand, understood intelligence as maturing from the inside, and being directed outwardly.

Again according to Sternberg (1993), more exciting than Vygotsky’s theory of internalization was his accompanying concept of the zone of proximal development. It is

true that one child can achieve greater developmental gains under the guidance of a particular teacher than can another. In essence then, the zone of proximal development is “the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers” (Sternberg, 1993, p. 210). Following this line of thought, there may be many high-ability learners who are not identified because their potential has yet to be acknowledged.

This concept has been popularized into a scenario which generally runs along the following lines. If one is given tasks that are consonant with one’s current ability level in any skill, then one is operating within one’s zone of comfort. Anticipation of the performance of such tasks engenders no particular anxiety within one; neither does one improve one’s skill by the performance of such tasks. Tasks that require skills that are far in excess of one’s comfort zone are so threatening that one avoids engaging with them, or fails summarily at them. Again little learning ensues from performing such tasks. Tasks that are so far below one’s comfort zone that one is able to give an automatic response to them engender a response of boredom. Again one tends to avoid such tasks, but if constrained to engage in them, one learns little. The crucial aspect for engendering learning is that the task should fall in the area slightly above our comfort zone; that is, when the task is beyond our current skill level, but not so far that we are reluctant to engage with it. In this case it is within our skill level but at the stage where performance of the task is challenging.

This concept of what Vygotsky referred to as the zone of proximal development has been found useful by many, notably Csikszentmihalyi--firstly in developing the concepts of flow and the autotelic personality (1991, especially p. 74), and then by him in

conjunction with Rathunde and Whalen (1993) in exploring the application of these concepts in the context of talented teenagers. Csikszentmihalyi et al. (1993) advert to the connection between the more recent Vygotskian formulations and Piaget's earlier conceptualizations of "the emergence of intelligence as the integration of two complementary processes—"accommodation" to outside reality and "assimilation" of what one learned from outside reality to mental schemes" (p. 79).

Second Movement §5(c): Mediation in dynamic vs static assessment

Feuerstein, who studied with Piaget, was squarely in the counter-Piagetian tradition of Vygotsky when he developed his concept of an assessment protocol which engaged a learner in a discourse as opposed to a monologue, or to use his terms, in developing a dynamic assessment tool as distinct from existing static tools. Feuerstein's basic premise is that intelligence is modifiable and that it develops by way of the mediated learning experience. This is

the way in which stimuli emitted by the environment are transformed by a "mediating" agent, usually a parent, sibling, or other caregiver. This mediating agent, guided by his intentions, culture and emotional investment, selects and organizes the world of stimuli for the child. The mediator selects stimuli that are most appropriate and then frames, filters, and schedules them; he determines the appearance or disappearance of certain stimuli and ignores others. Through this process of mediation, the cognitive structure of the child is affected. The child acquires behavior patterns and learning sets, which in turn become important ingredients of his capacity to become modified through direct exposure to stimuli. (Feuerstein, 1979, p.16)

Thus, it is more than acceptable for the assessor to interact with the child in an

assessment situation. Feuerstein adjusts the testing situation so that what is being evaluated is the child's ability to learn in the sense of the zone of proximal development, rather than what he or she already knows.

Kirschenbaum (1998) goes to some length to stress that the ability to distinguish between dynamic and static assessment should not be used to imply that the two types are not complementary: "dynamic assessment actually starts with a static measurement, but then allows the examiner to actively guide the student to the discovery of the solution through the use of "scaffolded" instruction" (p. 142). An important aspect for Feuerstein, developing out of the concept of the zone of proximal development, was that of "potential"—that idea that is nicely encapsulated in the coined concept of someone's "educability," or a measure of one's potential productivity outside testing situations (Sternberg, 1993). Lidz (1991) summarizes all the foregoing elegantly.

The ZPD concept refers to the idea that a child has some fully matured processes that are evident when the child is assessed by traditional means, as well as emergent developmental processes that can become evident when the child interacts with a more knowledgeable partner. The ZPD is the difference between the child's level of performance when functioning independently and the child's level of performance when functioning in collaboration with a more knowledgeable partner. This can also be viewed as a definition of "potential." (p. 7)

Clearly if one is able to implement some strategy, some procedure, that enables one to impute potential to some student from an underserved population, who possibly by virtue of straitened environmental circumstances alone, may be unable to demonstrate ability or achievement, then the path to correcting the bias of the current unfair practices

is clear.

Second Movement §5(d): Dynamic assessment, its derivatives, and the identification of high-ability learners

According to Passow and Frasier (1996), “dynamic assessment focuses on the specific behaviors, the ways the attributes are displayed in a particular context” (p. 201).

Kirschenbaum explains: “In dynamic assessment, the examiner provides scaffolded instruction that is either based on a standardized, hierarchic sequence of hints and prompts, or is more individualized, helping the student to complete the presented task, then records the effect of the assistance” (p. 142). In this way, “The goal of the assessment and intervention procedures is to help students develop cognitive skills commensurate with their true intellectual ability, not to increase their IQ scores or make them smarter than they would have been if they had an appropriate education” (Kirschenbaum, 1998, p. 142).

This approach has been adapted in a number of designs which Kirschenbaum classifies as dynamic assessment or dynamic-like assessment (1998). Instances of dynamic assessment include the Eureka model (Zorman, 1997), the mathematical task investigations of Jitendra and Kameenui (1993), and Borland and Wright’s (1994) use of dynamic assessment in identifying young, disadvantaged students as part of a gifted identification procedure. Instances of dynamic-like assessment include the Kay and Subotnik (1994) implementation in an arts program for inner-city, elementary school students, and Coleman’s (1994) use of dynamic assessment as an adjunct to what was essentially a portfolio assessment approach to examining a program for disadvantaged third graders.

Kirschenbaum (1998) states that “the advantage dynamic assessment has over

more objective, static assessment is that it is flexible enough to allow an examiner to explore ways of encouraging the demonstration of ability by helping a student to succeed at the task” (p. 144). This is in keeping with the assumption made by Vygotsky (1978) that a primary attribute of the developing human central nervous system is flexibility—“the ability to see alternative representations or adopt alternative strategies, especially when it is necessary to make a change for success on a task” (Shore & Kanevsky, 1993, p. 138). The operationalization of dynamic assessment faithfully implements the theoretical base from which it was developed.

Second Movement §6(a): High-ability learners

The term “high-ability learner” was operationally defined in the key terms section as referring to one who consistently learned how to perform largely novel tasks after minimal explanation. The concept of novelty was key in that for the student to be considered a high-ability learner, she or he was required to implement whatever internal processes were involved in learning, and to demonstrate by performance that learning had taken place. The items on the Project STAR test instruments were sufficiently in advance of what students at Primary and Intermediate developmental levels would be expected to know that the students would be operating in their zones of proximal development. The student who managed to adjust to the demands of the task and produced a high-level response would be adjudged a high-ability learner, and thereby be considered gifted under the South Carolina regulations.

Clearly if the target tasks for the learner are all pitched at too low a level, then the conclusion may be that everyone tested is a high-ability learner. Such a conclusion from non-discriminatory tasks is of little significance for the sample tested, but the same set of target tasks which were non-discriminatory with one sample may be appropriate with

another. The significance of the conclusion that one is a high-ability learner is predicated on the developmental appropriateness of the tasks as well as their novelty, and one's ability to perform them after one, or minimal, explanation. While this is somewhat less than Passow (1986) described years ago as an environment conducive to identifying talent among economically disadvantaged, racial/ethnic minorities, and limited English proficient students, the point of utilizing a modification of dynamic assessment for high-ability learners is precisely to create a micro-environment where it is "possible for students to engage in rich learning opportunities as a means of displaying gifted behaviors and talent potential" (Passow & Frasier, 1996).

Second Movement §6(b): Adapting dynamic assessment to fit the clientele

As mentioned above, Feuerstein and his colleagues devised a Learning Potential Assessment Device (LPAD) (Feuerstein, Rand, & Hoffman, 1979), but this again was specifically oriented to "retarded" individuals—in which context it uncovered talent in heart-warming ways (Kirschenbaum, 1998). What is being advocated in Project STAR is a modification of the technique itself, rather than an adaptation of a particular implementation. This study relates to the use of a dynamic-like assessment protocol as an adjunct to existing protocols in specific instances where it is particularly likely that the effects of socioeconomic disadvantage may have overly influenced the decision.

The Borland and Wright (1994) study mentioned earlier as an example of a dynamic-like assessment, provides an example or relevant implementation of some of the ideas advocated above in the environment of the underserved learner. In dynamic-like assessment the difficulties of understanding directions is obviated. Testers can pantomime solutions, or as in Raven's (1987) suggestion, repeat directions, or explain the task in detail, because the emphasis is on comparing the child's ability to perform the task

when assisted, to his/her ability to perform it alone. The child who may be puzzled about the task does not have to guess what is in the test administrator's head, or work out of an inadequate mental schema of the task. He or she can simply ask, and have any issues about the task clarified.

Second Movement §7: Dramatism

The time has come to consider in stark simplicity whether, in writing of high-ability learning this study is writing about something physiologically identifiable, or whether high-ability learning has to be understood as a hypothetical construct. If the conclusion at the end of this second movement is that high-ability learning is indeed a hypothetical construct, then there is good precedent for nonetheless proceeding. As Wertsch (1998) pointed out by reference to Dewey (1938, p. 263), hypothetical constructs are "inherently necessary for controlled enquiry." All of the preceding discussion has indicated that high-ability learning is at least a hypothetical construct.

Evidence that high-ability learning is more than a theoretical construct in sub-human species is starting to emerge from recent research, primarily that conducted on mice. In 1949, a Canadian psychologist, Donald O. Hebb

came up with a simple yet profound idea to explain how memory is represented and stored in the brain. In what is now known as Hebb's learning rule, he proposed that a memory is produced when two connected neurons are active simultaneously in a way that somehow strengthens the synapse, the site where the two nerves touch each other. At a synapse, information in the form of chemicals called neurotransmitters flows from the so-called presynaptic cell to one dubbed the postsynaptic cell. (Tsien, 2000, p. 63)

This strengthening of the synaptic transmission, or long-term potentiation (LTP)

has been shown to occur in reaction to high-frequency electrical impulse stimulation in the hippocampus—"a crucial brain structure for memory formation in both humans and animals" (Tsien, 2000, p. 63). This happens with the involvement of N-methyl-D-aspartate (NMDA) receptors—"miniature pores that most scientists think are made up of four protein subunits that control the entry of calcium ions into neurons" (Tien, 2000, p. 63). Conversely, low-frequency stimulation of the same pathways produces long-term depression (LTD) of the strength of the connection. Here then is a prime physiological mechanism for memory and forgetting at the atomic level. Merely finding such a prime candidate, however, does not prove that it is involved.

To move from circumstantial evidence to direct evidence, Tsien (2000) used genetic engineering techniques "to delete a sub-unit (the NR1 sub-unit) of the NMDA receptor in only a specific region of the brain" of mice. As expected, these mice "exhibit(ed) abnormal spatial representation and have poor spatial memory: They cannot remember their way around a water maze" (p. 64)). Later experiments showed that such mice demonstrated several other impairments in nonspatial memory tasks.

To explore even further the involvement of NMDA receptors in memory in mice, Tsien (2000) next bred mice with an extra copy of the gene which directs the production of the NR2B subunits of the NMDA receptor sites. Younger individuals in species as "diverse as birds, rodents, and primates" (p. 66) switch from making NR2B subunits to NR2A subunits as they mature. The major difference between the two subunits is that the NR2B subunits remain open for longer than NR2A subunits, thus increasing the likelihood of learning occurring as envisaged by Hebb (1949). Tsien, Liu, and Zhuo (Tsien, 2000) reported that such genetically engineered mice had NMDA receptors which stayed open for 230 milliseconds, roughly twice as long as those of normal mice.

These “Doogie” mice (named after the TV fictional character “Doogie Howser, M.D.”) performed at differentially superior levels, compared to normal mice, in tasks involving object memory, shock aversion, and the Morris water maze—a milky pond in which the location of a just-submerged rest platform is indicated by marks on the pond wall.

The purpose of reviewing this research in depth is that it demonstrates that there is a candidate physiological substrate for high-ability learning in sub-human species. The findings of Tsien (2000), his colleagues, and many other physiologists are beginning to show that the concept of high-ability learning may in the future become, to return to Dewey’s elegant phraseology, a “linguistic expression of something already known which needs symbols only for the purposes of convenient recall and communication” (Dewey, 1938, p. 263).

This physiological research has shown that performance-based identification based on high-ability learning may be able to tap into a very real substrate. To apply the sub-human analogy, it may well be that the children who are identified through performance-based instruments are those with a higher proportion of NR2B subunits than their colleagues. Of course, it is likely that the real situation is much more complex than any simple one-to-one correspondence such as that made here would suggest. If it was as straightforward as this, it is unlikely that the issue of the physiological substrate of learning has resisted full explication since at least Hebb’s (1949) day. And yet the thrust of Tsien’s (2000) research implicates the involvement of some underlying physiological structure in high-ability learning, which takes this concept beyond the purely hypothetical stage.

All of this exciting research has been occurring in a field not directly related to

education, yet its implications are far reaching for educators. These results are well-researched and stable enough to be spawning a number of pharmaceutical start-up companies seeking to apply the knowledge gained to date to alleviate the symptoms of Alzheimer's disease, for example. The implication of a future successful implementation would be the equivalent of a paradigm change for the field of education (see Kuhn, 1962).

Burke (1966) had much to say about what he referred to as the learned incapacities and disciplinary pathologies that restrict the horizons of modern academic discourse. The extended discussion of the light shed on this study by the ongoing study of learning and memory in animals is an illustration of the value of considering cross-disciplinary insights. The research so powerfully illustrating that high-ability learning has a definite genetic component in subhuman species brings this second movement nicely to a reprise of the first.

To stay with Burke's (1969) thought, he maintained, in common with Vygotsky (1978) and others, that "describing, interpreting, or explaining action, as opposed to some other phenomenon such as behavior, mental, or linguistic structure or attitudes" was of the utmost importance (Wertsch, 1998, p. 12). From this perspective, the insights gained from the physiological perspective above are valid, but when applied to the human level, are excessively reductionistic. Complexity is of the essence in these performance-based learning tasks. It is likely that the motivation for the action of the rat swimming to find the rest platform in the Morris maze is fairly uncomplicated. This "elementary" situation has few analogues in the everyday lives of most of us. Burke (1969) preferred to discuss human action and motives in terms of a pentad.

We shall use five terms as generating principle of our investigation. They are:

Act, Scene, Agent, Agency, Purpose. In a rounded statement about motives, you

must have some word that names the *act* (names what took place, in thought or deed), and another that names the *scene* (the background of the act, the situation in which it occurred); also, you must indicate what person or kind of person (*agent*) performed the act, what means or instruments he used (*agency*), and the *purpose*. (p. xv, italics and parentheses in original)

As Wertsch (1998) pointed out, there is a deceptive simplicity to Burke's pentad, yet in drawing this study to its conclusion, it is to this formulation that the discussion will return to gain perspective, and to look forward from the bold endeavor which has been and which is Project STAR.

Summary

There is ample evidence of ethnic imbalance in the identification of high-ability learners. Beyond the racial imbalance per se, there is the fact that students from economically disadvantaged families are less likely to be in special programs for high-ability learners (OERI, 1993). These imbalances have been adverted to by many researchers and various reasons have been advanced for the imbalance (e.g. Passow, 1989; Richert, 1987; Van Tassel-Baska, Patton, Prillaman, 1989). The two factors of ethnicity and socioeconomic disadvantage are often intertwined. NELS (1988) reported that only 9% of students in gifted and talented education programs were in the bottom quartile of family income, while 47% of program participants were from the top quartile in family income.

Considerable theory and some research exists to validate the use of alternative measures to uncover high-ability learning. Performing a high-level task may be one measure of learning potential. This theory has been set in the context of a dialectical format contrasting the positions of Vygotsky and Piaget—to mimic the dialectical nature

of the process of learning, which is something upon which they both agreed. The historical roots of this understanding were adverted to in passing.

Focusing on high-ability learning in the context of performing a problem-centered task may de-emphasize the advantages accruing to those students from more affluent families. If so, performance-based identification promises to be useful in ameliorating the outcomes of traditional aptitude and achievement tests. The reality that there are differing views about what the outcomes of traditional tests signify was intended to show that they are very useful for the purposes for which they were designed, but that they have characteristics which may well bias the outcomes when using them for gifted identification purposes.

The fact that there is a proportional imbalance among the subsections of the population from which come those currently served as high-ability learners points to the need for some adjustment. Passow and Frasier (1996) proposed the “ineffectiveness and inappropriateness of the identification and selection procedures” (p.198) was the most widely accepted explanation for the low participation of disadvantaged students in programs for the gifted. Alternative explanations of the imbalance include the possibility that those from the under-represented segments of the populations tend not to stay in the programs if they are identified, or a combination of these two factors and derivatives thereof (Borland & Wright, 1994).

Chapter Three

Methodology

Overview

This study was intended to contribute to the discussion of how to identify high-ability learners from socioeconomically disadvantaged backgrounds by launching a practical, easily administered, manipulative intensive, simply scored test which would be relatively culture-free in comparison with some of the frequently utilized forms of assessment for high-ability learners. The study concentrated on young, potentially gifted students from a variety of socioeconomic backgrounds.

Research questions and instruments

The initial question in this study concerned whether the instrument, specifically designed as a modified dynamic assessment instrument, had inherent credibility as a testing instrument. Hence the first research question is:

1. Do the Project STAR testing instruments exhibit reliability such that they can claim credibility as testing instruments in the task of identifying children to be given access to enhanced educational programming?

Credibility as a testing instrument involves more than reliability. To confront issues to do with content validity, reference will be made to the nature of the design process itself, the stages of review, and revision and refinement of items built-in to the process. The construct validity of the test will be addressed by exploring the psychometric properties of the instrument. The correlations between the Project STAR instruments and two traditional tests will be investigated. One of these will be a test of ability (Otis-Lennon School Ability Test, Sixth Edition—OLSAT), and one a test of achievement (Metropolitan Achievement Test, Version 7—MAT-7). These will be used to investigate the relevance of the Project STAR performance-based identification instruments to the characteristics traditionally held academically valuable.

The Cronbach alpha statistic will be used in considering the reliability per se of the instruments. This immediately raises the issue of the unit of analysis for this reliability calculation, and brings in a research sub-question:

- 1(a) Is there a basis for considering the Form A and Form B of the Project STAR instruments equivalent forms?

All of the above will be investigated on the heterogeneous-ability pilot study which preceded the field test from which the data for the remainder of the analyses will be drawn.

The second task of this study is to address the question:

2. Do the outcomes on the Project STAR instruments exhibit a bias on the basis of gender?

To address this issue, this study will turn to the field test sample, which consisted of students selected because they reached the criterion for identification as gifted in South Carolina in either Dimension A or Dimension B, but not both. The results on the two different forms will be anchored, using the results of the children on South Carolina's PACT test to develop a linear transform from one to the other in the process used to address research question 1 (a). A randomly selected sub-sample of half the children from the two combined levels (Primary and Intermediate) will be formed and exploratory factor analysis will be conducted. Once factors have been designated, confirmatory factor analysis will be used to examine the extent to which the exploratory factors are evidenced in the responses of the male and female children in each of three grade-level groups in the other half of the randomly selected subsample.

This same technique will next be used to investigate the question:

3. Do the outcomes on the Project STAR instruments exhibit a bias on the

basis of ethnicity?

As has been mentioned already, children whose families are from minority-language backgrounds are likely to be at some disadvantage when confronted with a traditional test situation. If the exploratory factors are just as relevant for African-American children as they are for the White children in the confirmatory phase, a strong argument for the unbiased nature of this performance-based instrument can be sustained.

Finally, the major concern of this study will be addressed:

4. Do the outcomes on the Project STAR instruments exhibit bias in terms of socioeconomic disadvantage?

The methodology will be the same as that used to address the preceding two questions. There is a hierarchy in these three questions in that gender indifference supports ethnic non-specificity, both of which support the final socioeconomic evenhandedness.

Research Design

There are two levels of analysis of interest to this study. At one level (question 1) this study draws on correlational research design (Gall, Borg, & Gall, 1996) on a heterogeneous ability sample. Hence, this study investigated the correlation between the outcome measures for subsets of the sample on the OLSAT, Mat-7 and Project STAR instruments. At another level, however, this study investigates a more far-reaching, conceptual issue inherent in the nature of the Project STAR instruments themselves (Crocker & Algina, 1984). At this general level, questions such as whether there were different latent traits underlying the responses of different subsets on the Project STAR instruments are paramount. Both levels of analysis are required to complete the picture. For example, there is little practical value in using an identification instrument if the

students it identifies have gifts in such abstruse fields that they are not likely to succeed in a school-based environment. Hence, some correlation with well-used measures of school ability is important. Similarly, basic adequacy as a testing instrument is paramount before any later analyses can be said to be indicative of anything. While not belittling these aspects of the forthcoming results of this study, the broader issues raised by the factorial analyses have been referred to above as far-reaching, and deserve the attention they will receive.

The quantitative measures on which this study focuses are not the only measures which are important in the identification of young, high-ability learners. Indeed, as has been made clear in the preceding, Project STAR implements Dimension C of the State Regulations, and in practice will only be used in South Carolina with students who have qualified on one of Dimension A or Dimension B, but not both. Rather than detracting from the significance of Project STAR, the fact that it is not being used as a first-resort instrument took some pressure off the development phase, but at the same time provided extra incentive to “get it right,” because for a number of children Project STAR could well prove to be their key to the door of effective learning. These expectations do not mandate a quantitative approach, but part of the rationale for Project STAR is the investigation of what promises to be just such a quantitative measure.

Site selection

This field study was carried out in South Carolina in twenty-eight school districts representing quite a range of demographic characteristics. All districts opted to be involved, and any district could withdraw at any time. Two districts which originally nominated withdrew before the field test began, leaving the twenty-eight included here.

Table 1

Sampling protocol showing systematic spreading of districts across forms

Size of district (# students)	Districts with $\leq 40\%$ minority student population		Districts with $> 40\%$ minority student population	
	Form A	Form B	Form A	Form B
≤ 5000	2	2	2	3
5,001 - 10,000	2	2	3	2
10,001 - 20,000	2	2	1	1
20,001 - 30,000	1	0	1	0
$> 30,000$	0	1	0	1
Total:	7	7	7	7

Note. From "Project STAR Field Test Administration Preliminary Report," by J. VanTassel-Baska, L.D. Avery, R.M. Reardon, and T.J. Ward, February 7, 2000, Unpublished report to the Project STAR South Carolina Steering Committee, p. 3.

Table 1 shows the way that the districts were split in terms of the number of students in the district, and the number of minority children in the student population. As no random assignment was attempted, a deliberate attempt was made to ensure an even coverage of type of school for each of the parallel test forms. If a district was assigned to Form A, and if that district was testing using both Primary and Intermediate level students, both levels answered Form A

Description of the test instruments

The Project STAR instruments at the field test stage consisted of an A and B form

pertaining to either Intermediate or Primary grade levels. The Intermediate level was designed to be used with children who were in either grade 4 or grade 5; the Primary level was designed to be used with students in either grade 2 or grade 3.

The Project STAR items were checked at the prototype stage with the South Carolina Standards to ensure alignment. Table 2 shows the topics from the South Carolina Standards document that were matched with each of the task prototypes at an early stage in the development of the items. (Because the Project STAR items were constructed under contract and are subject to the security requirements enforced for testing material in that state, copies of the actual items are not available for inclusion in this document. However Appendix A contains one verbal and one nonverbal item at the primary level, and one mathematical item at the intermediate level which were culled during the development process, for a variety of reasons. These items are indicative of the types of items actually included.)

The student books were printed on 17" x 11" paper which was folded to form standard 8.5" x 11" booklets, saddle-stitched on the spine. Each domain at each level was printed with different colored covers to provide visual cues for sorting. The inside pages were printed back-to-back, and care was taken, wherever feasible, to have all the prompt material visible on one page, or on the open two-page spread. This sometimes required blank pages, which were always clearly labeled as being intentionally left blank.

Table 2 shows the matching maintained between the South Carolina standards and the Project STAR items at the task development phase. This demonstrates the fact that content validity was planned into the task development process.

Table 2

Correspondence between Project STAR prototypes and the South Carolina Standards

	Task prototype topic	South Carolina Standard topics
Mathematics	Arithmetic problem solving	Arithmetic facts Problem solving Base 10 place values
	Number concepts	Ratios Factors and multiples Primes Whole number operations
	Logic	Mathematical reasoning
	Proportional reasoning	Exploration of ratio and proportion Formation of ratio
	Patterns	Recognition, extension, description, analysis of patterns
	Number theory	Connection of concepts in geometry and number Factors and multiples
	Spatial	Spatial problem solving
Patterning		Recognition and extension of patterns
Geometry		Perimeter Construction of geometric figures with concrete objects
Spatial reasoning		3-D models constructed from nets Identification of different views of a 3-D object
Transformations		Investigation and prediction of results of transformations Symmetry
Spatial visualization		Geometric patterns Thinking about geometric figures
Verbal	Verbal problem-solving	Explain author's purpose Make inferences from text Support fact and opinion with relevant details Analysis of literature
	Writing persuasive essay	Organization of writing Description of details Writing for an audience
	Vocabulary	Acquisition of rich vocabulary
	Analogies	Use knowledge of analogies
	Concept development	Use pictures to comprehend print materials Writing on a central idea
	Verbal reasoning	Use of evidence to support opinions
	Verbal relationship	Use of word-analysis skills

Note. From "Project STAR Development of Student Task Assessments and Rubrics. Year 1 Report: Pilot Phase," by J. VanTassel-Baska, T. Ward, D. Johnson, L. Avery, and L. Dolins, 1999, Unpublished report to the Project STAR South Carolina Steering Committee, p. 14.

The Intermediate level tests consisted of three domains: verbal, mathematical, and

spatial at the field test stage, and the Primary level tests consisted of two domains: verbal, and nonverbal. Each domain contained five test items. The nonverbal domain at the Primary level was formed by amalgamating the mathematical and spatial domains in the process of review after the pilot test. The major motivation for this reduction was twofold: feedback from the schools which indicated that the instrument was too long for the younger children, and psychometrically the additional items were not contributing sufficiently to reliability to retain their use. Hence, in the course of the review, three of the mathematics items and two of the spatial items which contributed least to the Cronbach alpha, and which appeared to be expendable also from a content analysis point of view were dropped. The remaining two mathematics and three spatial items constituted the new nonverbal domain. This process of item review and subsequent culling had been established prior to the pilot testing phase at which time the number of items per domain had been reduced from six to five.

Basic test protocol

Each of the items contained a preteaching example which required the teacher to demonstrate a particular skill. The children were encouraged to ask questions. When the teacher was satisfied that the children could carry out the process involved in the preteaching, he or she instructed the children to open the student book, read the item with the children (the item always drew heavily on the preteaching; the older children were given time to read rather than having the teacher read the item for them), and gave the children fifteen minutes of solo work-time to complete the task. The teacher rendered no assistance from the time the students mastered the preteaching example. The fifteen minute task time was intended to ensure that the Project STAR tasks were not speeded, and such has proved to be the case in practice. The allotted time was found to be ample to

complete the tasks, though it was reported from a few test sites that some students became so absorbed in a particular task that they were reluctant to move to the next. The teacher was empowered to exercise judgement and move on to the next item if all the children seemed to have produced as much as they were likely to produce. Breaks during the course of the test were allowed, at the teacher's discretion. Teachers were asked to record the number and the duration of breaks. It was strongly recommended to the teachers that all the items in any one domain be completed in one block in the day.

Training requirements

Test administrator's training

Representatives from every district involved in the field test assembled for one of two day-long training days prior to the commencement of the testing. These training sessions were conducted by a staff member of the Center for Gifted Education. The intention was that a high level of familiarity with the performance-based tasks on the part of the administrators would lead to closely similar testing conditions. These test administrators were introduced to the concept of performance-based assessment and given something of its background. They were then walked through every item in the test they were to administer, and the directions already printed in the teacher's book were explained and, if necessary, interpreted. Administrators were supplied with professionally produced overhead transparencies, and sets of manipulatives packets to aid in the pre-teaching modules. Some districts intended using several testing sites, in which case the person attending carried out a training session for the other test administrators in the district. Test administrators took all the required booklets and manipulative packets with them when they left the training.

Test scorer's training

The student booklets were all scored in Columbia, South Carolina. Test scorers were trained “on the job.” The basic design called for scorers to work in dyads. The two people initially worked together to score one booklet. Issues that arose were discussed with the trainer before the pair began working separately—initially with frequent review and close oversight. As the trainer and the scorers gained confidence in the scoring rubric, the trainer withdrew from first-hand contact with a scoring dyad. Although one person scored each booklet (there was a separate booklet for each domain), the second member of the dyad was always easily accessible for consultation. In the event of a disagreement, one of the two members of the state steering committee who were in attendance at each scoring session, adjudicated the outcome. These arbiters also checked scored booklets on an ad hoc basis.

To help to standardize the scoring task, a scoring rubric booklet was produced. This not only contained a synopsis of each student task, but included grading instructions, and exemplars of expected responses. For most items, a raw score was produced by totaling the relevant aspects of the student’s response and converting this raw score to a rubric score. Both raw scores and rubric scores were recorded on the front of the scoring booklet and in the data management system, though only rubric scores were used for analysis.

Other requirements

One member of the State Steering Committee assumed the task of creating packets of manipulatives for each teacher and each child. This required the assembling of approximately two thousand packets of manipulatives for an average of six items on each of four level/form combinations. This was no small task, and with only minor hitches

amounting to some poorly cut patterns on one particular item.

Participants

South Carolina State Regulations designated three dimensions on which a child could show high ability should they fail to qualify for identification as academically gifted on the basis of outstanding excellence. Dimension A concerned a child's reasoning ability. A student was considered as potentially eligible for special services if he/she ranked at the 90th percentile or higher on a national age percentile basis. In the Project STAR field test, the pre-eminent instrument of choice was the Otis-Lennon School Ability Test (OLSAT). Dimension B concerned a child's achievement, and here a student was considered to be potentially eligible for special services if he/she ranked at the 94th national percentile or higher in reading and/or mathematical areas. In the Project STAR field test, the overwhelming instrument of choice was the Metropolitan Achievement Test, 7th edition (MAT-7).

If a student is ranked in the indicated percentile range on both Dimension A and Dimension B, he/she was declared to be eligible for special services. If the child was eligible in either Dimension A or Dimension B, but not both, he/she was then eligible to be considered under Dimension C, the newly developed Project STAR protocol.

With this as background, this study was concerned specifically with the outcome of utilizing performance-based tasks with the sample of young students (i.e. grades 3-6) who qualified on either, but not both, Dimension A or Dimension B of the South Carolina regulations governing the identification of academically gifted children. This group of students is in an invidious position in that they are acknowledged as being high functioning in one of these Dimensions, but at the same time are not eligible for specific educational programs which may trigger their potential because they are just below the

cutoff on the other. There are quite a number of students who fall into this category, 1792 of whom participated in the field test.

Each participating school district was requested to designate up to twenty five children in each of grades 3, 4, 5 and 6 who met the State criteria for identification as above on one of Dimension A or B but not both. In practice, few of the school districts succeeded in designating exactly twenty five students in the required grades. This was due to various reasons. Some districts were too small to have twenty five such children, while in others there were some schools where the principal implemented a policy which prevented the testing of students who were not already identified, thus defeating the point of implementing Dimension C.

In summary, the participants represented convenience-sampled groups of students (Grade 3: N = 478, Grade 4: N = 483, Grade 5: N = 435, and Grade 6: N = 372. Total N = 1768) from twenty eight, convenience-sampled school districts which were ranked and then paired on the variables of size of district, expenditure on educational resources, and ethnic proportions (see Table 1 for sampling protocol).

Other instrumentation

Test administrators were requested to provide a report of the testing event, noting especially any unusual conditions that may have impacted on the students' ability to engage in the tasks. If the test administrator felt it necessary to issue directions not included in the administration booklet, he/she was requested to report the directions given, and the issue which occasioned this step being taken. Test administrators were asked to report the day and time of day the test was held. Administrators who had to schedule the administration of the test on two different days, or at different times in the same day, or who allowed breaks, were asked to provide the detailed schedule of these.

Data collection

Scorers recorded their decisions directly on the front cover of the student books in a scoring matrix printed there. The front covers were then detached and forwarded to the Center for Gifted Education for data entry. The complete booklets of four districts were preserved and returned to be used to update the scoring exemplars, should this prove necessary.

Demographic and standardized testing data had been entered into a special purpose database prior to the arrival of the Project STAR results. The booklets were bundled by district and color (which corresponded to domain, as mentioned above), and then entered into a special-purpose database, matching the already entered data on the basis of the student's identification number (which in South Carolina is the same as the child's social security number). Upon completion of a district's data entry, a formatted output report was printed and returned to the Project STAR contact in that district for checking. In this way a number of errors were notified and corrected, and a revised printout was returned to the Project STAR contact. This iterative step also resulted in the supply of some of the data missing at the initial entry stage, and contributed significantly to the establishment of a clean data file.

Statistical Procedures

The first question that needs to be addressed with these instruments relates to whether they could be reasonably described as psychometrically robust. In order to address this issue, the student results on each of the forms of the instruments were examined in terms of reliability, which was defined in terms of the Cronbach alpha coefficient. This is a measure which can be characterized as the average of all possible split-half reliability coefficients, and as such is a good measure of internal consistency.

The student results were correlated with the student's prior results on the OLSAT and MAT-7 tests. The purpose here was to achieve a somewhat low but significant correlation, in the range of 0.3 to 0.6. A correlation of this order would signify that these instruments are in fact testing something different from those which produce the Dimension A and Dimension B results—a situation that is clearly highly desirable. At the same time, however, a significant correlation is desirable because this will signify that this instrument is not testing something which could be characterized as not relevant to general educational goals.

It is desirable that any instrument to be used with such a select group of students will be effective in spreading the participants along a new axis, as implied by such correlations as were discussed in the previous paragraph. Widely used ability and achievement tests, like OLSAT and MAT-7 typically choose to employ a value for the standard deviation of the standardized scores of about one fifth to one sixth of the mean value. Given that this Project STAR analysis yields standard deviations in accord with this expectation, it is then relevant to ask whether the factorial structure as revealed in the participants' responses corresponds to the domain descriptors which were the a priori bases for item generation, and particularly whether it is the same for the major sub-samples on the basis of gender, ethnicity, and especially socioeconomic disadvantage. Hence a thorough investigation of the factor-analytic deconstruction of the participants' responses will be carried out.

The final three families of analyses will have to do with a quite crucial aspect of this instrument which has to do with the issue of bias. In this tightly defined population, the distribution and even the existence of high-ability learning is unknown. Given this, it is nonetheless desirable that the outcome for children in one particular sub-sample of the

population will be similar, in terms of the latent factors identified, to the outcome for children in the complementary sub-sample. The three major sub-sample dichotomies which will be investigated are: gender, ethnicity, and socioeconomic disadvantage.

Time frame for this study

The project-based data for this study was collected in fall 1999.² Analysis for this study proceeded in concert with the analysis for Project STAR, but pursued extended lines of inquiry. Hence, while the psychometric aspects of the instruments were relevant to both the Project STAR report and this study, all of the factor analysis, both exploratory and confirmatory, were conducted for this study only. It is anticipated that the analysis of the data for this study will be completed by the end of April, 2000.

Limitations and delimitations

One of the limitations of this study was an outcome of the time-line inherent in its application. This restricted the extent to which items could be tried out with children of the target age. Items were tested locally with about 600 students during a two-week slot in March 1999. These local, mini-tryouts enabled a sense of the timing and difficulty level to be refined, and they provided an opportunity for the scoring rubrics to be tested on actual responses. A tryout phase proper was conducted in South Carolina with small groups of children in grades 2 and 5 in five districts. The student responses were scored

²

Extremely tight deadlines had to be adhered to in order to fulfil the South Carolina Department of Education's need to have results from the field test to inform the State Steering Committee's decision to proceed to statewide implementation in spring 2000. Data entry by multiple people at the Center in November 1999 was supported by the custom-built, multi-user database, and all data were entered before the Christmas 1999 break. The analyses required to enlighten the Department of Education's decision were reported at the end of the first week of February 2000. Those analyses address only the first research question presented in this study, in which issues of validity and reliability of the Project STAR instrument were considered.

by members of the task development committee and teachers in Columbia towards the end of March 1999. Although the outcomes of both these try-out phases were carefully combed for implications for the items, there was insufficient time to re-try the revised items, and to try-out new items building on what had been uncovered prior to the pilot test.

Another set of limitations had to do with the testing protocol itself. There is an inherent lack of control over a number of variables in the testing protocol. For instance, there was some evidence of over-enthusiastic teacher/testers being too intrusive in the testing process. This was surmised when the five responses from one small testing center all incorporated the same (erroneous) wording. Teacher/testers had been instructed when to change from teacher-mode to tester-mode, but this was arguably a case in which this did not occur. Another variable over which there was little control was the length of break given during the testing process. Because of the length of testing time involved (teaching time plus an average of 15 minutes response time) provision was made for a bathroom break at about the mid-point of the testing. Teacher/testers were asked to record and report on the length of the break, and to limit its duration, but this was still an aspect that was largely uncontrolled.

A third set of limitations had to do with the implementation of the scoring rubric itself. There were questions about the interpretation of some responses which were solved by running rules as they were brought to the scoring supervisor's attention. Furthermore, scorers had to succeed in a training session before actually scoring any student work, scorers worked in league with a consultative partner, and the scoring decisions of every scorer were checked on a regular basis. However, the reality was that some variability was evident in scorers' interpretation. This variation amounted to no more than one rubric

point in any individual case, but deviation from the rubric to even this small extent was enough to potentially create difficulties for some students in reaching criterial levels. Sometimes the error amounted to a mis-transcription from the raw score to the rubric score, which raised the issue of whether the raw score itself may have been the better score to record. The raw score could have easily been transformed to a rubric after entry into the computer.

Along this same line of thought, the rubric data was entered by hand by a group of graduate students at the Center for Gifted Education, leading to the possibility of data entry errors. The data entry was checked for errors by perusing scatterplots of the data for anomalies, and many records were spot checked. Some errors were detected by both methods and corrected. The data entry error likelihood has been lessened in the current iteration of Project STAR by using optical mark sense sheets for the scorers to record their scores.

In summary, the limitations on this study could be attributed to the time-line inherent in the implementation, and to the performance-based, classroom-administered nature of Project STAR, and the fact that it is not a standard psychometric test. These limitations have been addressed and where possible, steps have been taken to ameliorate their effects.

Two delimiting factors connected with this study were the narrowing of the scope of the investigation to the three grades for which PACT data were available, and the restricted window on the panorama of dynamic assessment to which the students were given access. The restriction of the study to grades 4, 5, and 6 was unfortunately part of the need to utilize the anchoring process as discussed at length in this study. The restricted exposure to dynamic assessment was directly related to the length of time it

took to conduct the testing session. Clearly there were far more domains to which this technique could have been applied, and, as discussed in this paper, the original spatial and mathematical domains were combined to reduce the testing load by removing items which contributed little to reliability and seemed to be overlapping other items in terms of content. While the delimiting factors were unwelcome, it is arguable that they may have affected the outcome.

Chapter Four

Results

Overview

This chapter reports the results of the four research questions of this study. The first question concerned the reliability of the Project STAR instruments. Many aspects related to validity have been dealt with in explicating the design process and the rigorous scrutiny which items had to survive in order to be selected to go forward from the design to the implementation stages of Project STAR. Reliability is a necessary, but not sufficient, condition for validity. The detailed reliability analysis reported here was carried out on the pilot phase of Project STAR, as discussed below, and completed the picture of the Project STAR instrument as a credible testing instrument. The question of anchoring of the two forms of the Project STAR instruments is answered in the affirmative, and the process is incorporated into the larger picture.

Having established the credentials of the instrument, the ensuing three questions which were presented above in increasing order of importance for this study were approached using the field test data. The first step in this set of analyses was to anchor the two forms (Form A and Form B) of the test at each level (Primary and Intermediate). Following this anchoring process, exploratory factor analysis of the results for a random sample of half the students was used to develop factors. The outcome of the exploratory factor analysis at each grade level then formed the basis for confirmatory factor analysis in the form of structural equation modeling, carried out on the half of the students not already included in the exploratory phase. The way in which the second half of the sample at each grade level was dichotomized determined the conclusion which was drawn. Dichotomies were developed along the lines of gender, ethnicity, and finally socioeconomic status.

All analyses were conducted using SPSS for Windows, Release 10.0.5 (27

November, 1999). The data were maintained in relational schema in FileMaker Pro 4.1, and exported in DBase III format to SPSS. Mathcad 8.0 was used to develop the linear transform equations involved in the anchoring process, after the necessary variables were calculated in SPSS. Amos, Version 3.62 was used to perform the structural equation modeling at the confirmatory factor analysis stage.

Reliability

The concept of reliability grew out of a theoretical model for “characterizing the influence of random errors on test scores” (Crocker & Algina, 1986). The reliability coefficient is equivalent to “the proportion of the observed score variance that is attributable to variance in examinee’s true scores” (Crocker & Algina, 1986). This concept is particularly apposite when using Cronbach’s alpha (one of a set of three measures yielding identical results—collectively called “coefficient alpha” procedures—developed in the 1930s and 1940s) as the measure of reliability. In the context of a single administration such as Project STAR, each item is interpreted as a subtest, and the total score for each domain is regarded as the composite.

The coefficient of reliability may be adversely affected if the results exhibit a restricted range of variability (Crocker & Algina, 1986). Hence it is acceptable practice to calculate reliability coefficients on heterogeneous samples (H. Huynh, personal communication, March 20, 2000). In this context, the pilot phase of Project STAR utilized a heterogeneous sample, so it was appropriate to calculate the reliability coefficients on the basis of that phase—in contrast to the later field test phase which utilized a sample of students who were eligible for Dimension C testing. The remaining questions were addressed in terms of the field test sample. Before proceeding to attend to research question 1, question 1(a) must be addressed:

1(a) Is there a basis for considering the Form A and Form B of the Project STAR instruments equivalent forms?

Crocker and Algina (1986) discuss just such a situation as was confronted here. They nominate a process of linear equating as the solution to the following case: two instruments which are to be equated are administered, each to a different group of examinees, and a single anchor test is administered to both groups. In the Project STAR pilot test, by design, no student did both Form A and Form B of the test. However all students in South Carolina do sit for a state-wide test of achievement, which thus becomes a candidate for use as an anchor test. Out of the anchoring process will come, in this case, a decision as to what is a reasonable basis for developing the coefficient alpha calculations.

Linear equating was appropriate in this case and the procedures for Design C (Crocker & Algina, 1986, p. 460) were conducted, utilizing the South Carolina Palmetto Achievement Challenge Test (PACT), taken by students in grades 3, 4, and 5 (among others) in spring 1999, as the anchor. According to Crocker and Algina (1986) the groups in Design C are not necessarily formed by random assignment. This anchoring process using PACT data was preferred to the use of the results on either the Otis-Lennon School Ability Test (OLSAT) or Metropolitan Achievement Test Version 7 (MAT7), both of which were also available (H. Huynh, personal communication, March 8, 2000). One possible effect of this decision is discussed later in this study. The PACT results were downloaded by courtesy of the South Carolina Department of Education and matched to the children using the child's social security number as the match field, wherever possible. Any incorrect length social security numbers in the PACT file were padded to the correct length with zeros, on the presumption (later verified) that the export

routine utilized at the South Carolina State Department of Education regarded leading zeros in a numeric field as non-significant. The social security number field in the PACT was then re-defined as text and the dashes inserted after the first three and medial two digits to conform to the usual social security number format used in the pilot database. Unfortunately in the pilot design, a number of social security numbers were either not recorded on the children's booklets, or recorded incorrectly. In these cases, a match field was generated by concatenating the first name, last name and grade of the child, and visually checking the potentially matching data in the PACT file. The results of pilot students for whom there were no PACT data at this stage were removed from the anchoring process. This step eliminated 184 of the 1425 records in the pilot file. A further 303 records were eliminated from the anchoring process because there were no PACT data available for grade 2 students, and a further 17 records for whom no PACT data were supplied were also removed, leaving a file 921 records. The final reduction in the size of the file eliminated those for whom a complete STAR data set was not available, leaving 824 records. The breakdown of these students across the grade levels and forms is shown in Table 3.

Table 3

Breakdown of pilot sample, showing numbers of students for whom PACT data was obtained

Level	Grade	Form A	Form B
Intermediate	5	153	121
	4	156	141
Primary	3	135	118

To recap, the pilot data is being used here to establish the reliability of the Project STAR instruments because the pilot test was conducted on a heterogeneous sample. It is acknowledged practice to use a heterogeneous sample to establish reliability because the restriction of range can lead to spurious reliability figures. Each of the three different grade levels (3, 4 and 5) must be anchored separately (H. Huynh, personal communication, March 20, 2000). The statistics required in the linear transform (developed in Mathcad) from Form A scores to Form B equivalents at each grade level are shown in Table 4. Once the B-equivalent scores are developed by calculation from the Form A scores in SPSS, the independent sample t-test was used to judge whether the Forms of the test were in fact equivalent and could be considered as part of the same continuum, or whether they were dissimilar and had to be considered separately.

In Table 4, the subscript numbers refer to the grade level, and the subscript capital letters refer to the two forms: A and B. The “M” represents the mean and “S” the standard deviation. The “b” represents the slope of the regression line of the subscripted group on the PACT data for that group. Hence “ $b_{3APACTA}$ ” refers to the slope of the regression line of the third grade STAR results on the third grade PACT outcomes for those who took Form A of the STAR instrument.

Table 4

Descriptive statistics for equating Form A results to Form B equivalents

Grade	Group	Statistic	Variable pertaining to:		
			Form A	Form B	PACT
3	A ₃	M _{3A}	28.49		618.16
		S _{3A}	7.54		29.50
		b _{3APACTA}	.212		
	B ₃	M _{3B}		27.03	605.87
		S _{3B}		9.09	27.82
		b _{3BPACTB}		.226	
	Total ₃	M ₃			612.36
		S ₃			29.31
4	A ₄	M _{4A}	34.25		817.01
		S _{4A}	10.97		31.47
		b _{4APACTA}	.298		
	B ₄	M _{4B}		34.04	812.67
		S _{4B}		10.51	30.80
		b _{4BPACTB}		.282	
	Total ₄	M ₄			814.94
		S ₄			31.17
5	A ₅	M _{5A}	38.28		1017.57
		S _{5A}	11.03		28.27
		b _{5APACTA}	.327		
	B ₅	M _{5B}		37.27	1010.83
		S _{5B}		13.43	30.47
		b _{5BPACTB}		.366	
	Total ₅	M ₅			1014.58
		S ₅			29.41

Before proceeding to use the statistics from Table 3 to produce the linear transform, it is important to note how an issue which arose in the course of developing Table 4 was handled. In producing the slope values, it became clear that one further adjustment was necessary to the sample for this analysis. The presence of clear outliers, corresponding in this case to students who scored very poorly on PACT and quite well on Project STAR, was unduly affecting the statistical values. This is illustrated for the grade 4 Intermediate Form A results as shown in Fig. 5. While the presence of outcomes of such divergence could be validation of the existence of a group of students to whom Project STAR is specifically oriented, it is nonetheless important to remove these results from the sample for the purpose of this exercise, because of the strength of the effect of such extreme outliers on the parametric statistics involved here. Figure 5 shows a group of three students who scored at the 400 level on PACT, but close to the mean on Project STAR. The extent to which they skew the statistics is illustrated by comparison with the same sample in Figure 6 with the outliers removed.

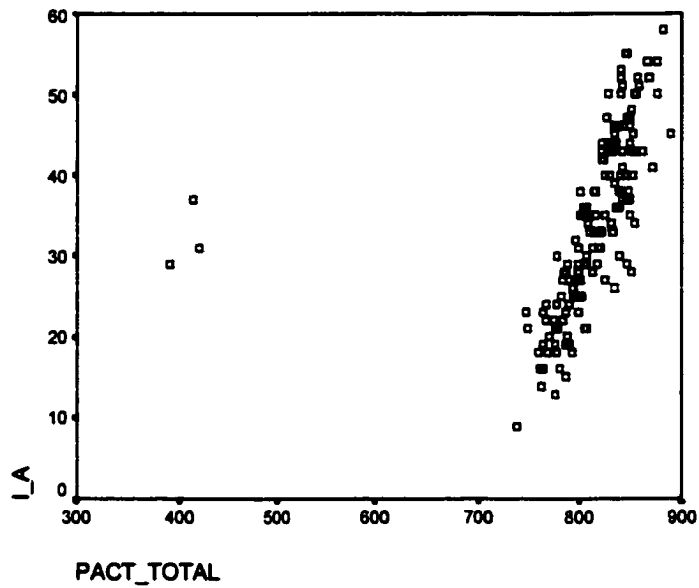


Figure 5 Grade 4 Intermediate Form A group STAR regressed on PACT, showing outliers on PACT results.

The effect of removal of the outliers is clearly seen in comparing Figure 5 with Figure 6.

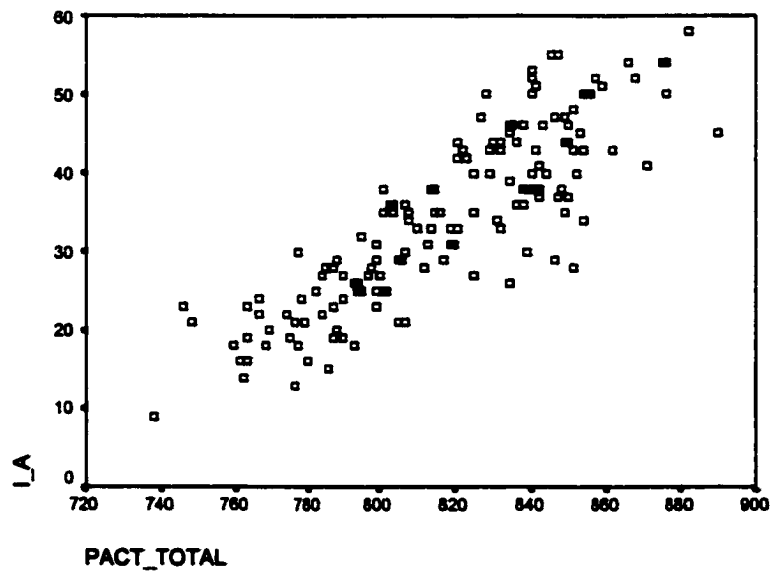


Figure 6 Grade 4 Intermediate Form A STAR regressed on PACT, showing the effect of removing the three PACT outliers in Fig. 5.

Having removed the outliers at each grade level, the final breakdown of the pilot sample ($N = 816$) is shown in Table 5. Table 4 above was adjusted subsequent to the removal of outliers and reflects the statistics for this Table 5 breakdown.

Table 5

Final breakdown of pilot sample, after removal of outliers

Level	Grade	Form A	Form B
Intermediate	5	152	121
	4	153	140
Primary	3	132	118

The basic equation for the linear transform from Form A scores to Form B equivalent scores (B^*) is:

$$B^* = a(A - c) + d$$

In this case, using the grade 3 statistics in Table 4 as an example,

$$a = \sqrt{\frac{S_{3BFormB}^2 + b_{3BPACTB}^2 (S_3^2 - S_{3BPACT}^2)}{S_{3AFormA}^2 + b_{3APACTA}^2 (S_3^2 - S_{3PACT}^2)}}$$

$$c = M_{3AFormA} + b_{3APACTA} (M_3 - M_{3PACT})$$

$$d = M_{3BFormB} + b_{3BPACTB} (M_3 - M_{3PACT})$$

The key to understanding these formulae is to read “c” as using the relationship between Form A and PACT to yield an estimate of the mean Form A score for the whole

group, and “d” as using the relationship between Form B and PACT to yield an estimate of the mean Form B score for the whole group. The numerator and denominator of “a” represent estimates of the variances of the Form B and Form A scores respectively (Crocker & Algina, 1986). Substituting the values for the appropriate statistics into the linear transform (all statistics correct to 2 decimal places) for each of the grades 3 through 5 leads to:

$$\text{For grade 3: } B^* = 1.24 (A - 27.26) + 28.50$$

$$\text{For grade 4: } B^* = 0.97 (A - 33.63) + 34.68$$

$$\text{For grade 5: } B^* = 1.16 (A - 37.30) + 38.64$$

These formulae were then used to generate the B-equivalent scores (B^*) for all those who took Form A and the difference between the Form B^* and Form B groups assessed using a t-test for independent samples. With an alpha level of .05 and a two-tailed test, the mean of the Primary students (grade 3 only) for the B^* data from the Form A group ($M = 27.0$, $SD = 9.1$) was significantly less than the mean for the B data from the Form B group ($M = 30.0$, $SD = 9.35$), $t(248) = -2.56$, $p < .05$. In contrast, for the Intermediate grade 4 group, with an alpha level of .05 and a two-tailed test, the mean of these students for the B^* data from the Form A group ($M = 35.3$, $SD = 10.6$), was not significantly different from the mean of B data from the Form B group ($M = 34.0$, $SD = 10.5$), $t(291) = +1.01$, $p > .05$. Similarly, for the Intermediate grade 5 group, with an alpha level of .05 and a two-tailed test, the mean of these students for the B^* data for the Form A group ($M = 39.8$, $SD = 12.8$) was not significantly different from the mean of the B data from the Form B group ($M = 37.3$, $SD = 13.4$), $t(271) = +1.57$, $p > .05$. These results are shown in Table 6.

Table 6

Summary of results for test of difference between B-equivalent (B*) and B scores after anchoring

Grade	B* mean (standard. deviation)	B mean (standard deviation)	t (df)	prob. (p)
3	27.0 (9.1)	30.0 (9.35)	-2.56 (248)	< 0.05
4	35.3 (10.6)	34.0 (10.5)	1.01 (291)	>0.05
5	39.8 (12.8)	37.3 (13.4)	1.57 (271)	>0.05

These results indicate that at the Primary level, the Form A and Form B versions yielded outcomes which should be dealt with separately, but that the Form A and Form B Intermediate versions at both grade 4 and grade 5 levels functioned as parallel tests. This, then is the ultimate answer to research question 1 (a): yes, it is possible to discern whether Form A and Form B have acted as parallel forms, and it is important to do so, because they do not automatically function in this way. Now, to direct attention to research question 1.

1. Do the Project STAR testing instruments exhibit reliability such that they can claim credibility as testing instruments in the task of identifying children to be given access to enhanced educational programming?

In terms of the reliability calculations, a field test-equivalent form of the pilot outcomes was formed by deleting from the calculation the one item that was removed from each of the dimensions to shorten the administration time at the field test stage (H. Huynh, personal communication, March 20, 2000). The above results on the anchoring process indicated that the reliability of the Primary Form A and Form B should be calculated separately, while the reliability for the Intermediate forms could be calculated

on the combination outcomes on the individual forms. The coefficient alpha reliability outcomes are as shown in Table 7.

Table 7

Coefficient alpha reliability of pilot "field test equivalent" Project STAR instruments

Level	Domain	Form A	Form B	Combined
Primary	Verbal	.76	.78	
	Nonverbal	.72	.77	
Intermediate	Verbal			.75
	Mathematical			.71
	Spatial			.77

These coefficient alpha reliability values are quite consistent across all combinations of level and domain, but would be regarded as being up to one decimal point below the usual range of acceptability for research instruments used for most purposes, commonly nominated as 0.80 or above (Gall, Borg & Gall, 1996). As mentioned in the introduction, reliability is a pre-condition for validity, and at first encounter, these lower values appear to be a disappointing outcome. One of the characteristics that would be expected to lessen the reliability of the Project STAR instrument is the number of items in the domains. Hambleton (1990) overviews the "large body of literature" (p. 402) on this issue in relation to criterion-referenced tests. Hambleton (1990) introduced the concept of precision to assist in determining the appropriate length of a test, based on the desired domain score distribution. In this application, there is no a priori domain score distribution, so it is not appropriate to calculate a value for the required number of items (Eignor & Hambleton, 1979). Nevertheless it is of interest to note that in Hambleton's (1990) example nine items were

required to yield even a modest degree of decision accuracy: “relation between decisions based on a test and decisions evolving from a criterion measure such as teacher ratings” (Hambleton, 1990, p. 402).

Whatever the etiology, the force of Project STAR’s lower-than-usually-acceptable reliabilities is further mitigated by consideration of Croker and Algina (1986), where, in discussing the effect of true score variance on reliability, they comment that “reliability coefficients ... have limited usefulness in assessing the quality of information provided by a test used for screening or selection” (p. 145). In these cases, Croker and Algina (1986) go on to point out, the issue is only “whether the examinees score is above or below a certain cutoff score” (p. 145). In this case “the magnitudes of the true and observed score variances (and their ratio) have less relevance for this measurement process” (p. 145, parentheses in original). Thus, while the coefficient alpha values are less than would usually be considered acceptable for a research instrument, the specific issue which the Project STAR instrument addresses—namely the dichotomous designation of students as high-ability learners—considerably lessens the deleterious effect of these lower values.

Before summarizing this aspect of the investigation, it is appropriate to draw attention to a “common misinterpretation” (Crocker & Algina, 1986, p.142) first clarified by Cronbach (1951, cited in Crocker & Algina, 1986, p. 142), namely that a relatively high coefficient alpha signifies a unidimensional test. As Crocker and Algina (1986) went on to explain, “because alpha is a function of item covariances, and high covariance between items can be the result of more than one common factor, alpha should not be interpreted as a measure of the test’s unidimensionality” (p. 142).

Summary of inquiry into reliability

The above investigation of the technical adequacy of the Project STAR instruments in terms of reliability has led to the conclusion that the instruments are adequate for the specific purpose for which they are intended. This less than resounding endorsement arises from the magnitude of the reliability coefficients. There are reasons for these coefficients to be less than one would usually expect from a test, as outlined above, and these should be read as mitigating circumstances. What should also be kept in mind is that the Project STAR instruments are not intended to be instruments of first recourse. Any child who is tested using the Project STAR instruments in practice will already have taken two other tests (typically the OLSAT and MAT-7) and so Project STAR will typically be asked to decide in cases similar to the ones which were pointed out as outliers in Figure 5 above. This sort of discrimination is clearly possible with the Project STAR instruments. A brief digression to consider how Project STAR fits in with OLSAT and MAT-7 follows.

Correlation between Project STAR and other instruments

While this correlational issue was not a major one for this study, it is useful to consider as a part of the technical adequacy related to the question of test validity. The findings here also act as a prelude to the factor analytic results which follow.

The Spearman correlation between the composite Project STAR scores and the OLSAT percentiles was 0.377 for Intermediate Form A and 0.286 for Intermediate Form B. Both of these were significant at the .01 level. For Primary Form A the Spearman correlation was 0.437, and for Form B 0.336; again both significant at .01 level. These low, but significant correlations were precisely what was desired for Project STAR. It was hoped that results on Project STAR would exhibit sufficient correlation to be

plausibly identifying something relevant to the schooling milieu, in the sense that OLSAT does. This is the interpretation of the low, but significant correlations which were found. The other positive outcome of this level of correlation is that it is clear that the Project STAR instruments are tapping into something quite different from OLSAT.

Similarly, for the Spearman correlations between Project STAR and the MAT-7 percentiles. The highest Spearman correlation between Project STAR and MAT-7 was between Intermediate Verbal (STAR) and Total Reading (MAT-7) at 0.426 ($p < .01$). The lowest was .017 for Project STAR Primary Nonverbal and MAT-7 Mathematical Concepts. While this latter correlation is clearly not significant, there is also little overlap in intention between the Project STAR Nonverbal and MAT-7 Mathematical Concepts. The conclusion reached was the same here as for the OLSAT above. The correlations showed that Project STAR was identifying something relevant to the schooling milieu, but clearly something quite distinct from what MAT-7 was measuring.

Factor Analytic Questions

2. **Do the outcomes on the Project STAR instruments exhibit a bias on the basis of gender?**
3. **Do the outcomes on the Project STAR instruments exhibit a bias on the basis of ethnicity?**
4. **Do the outcomes on the Project STAR instruments exhibit a bias on the basis of socioeconomic status?**

These questions represent the main thrust of this study, since they cover the issue of bias in the Project STAR instruments. Before proceeding to consider these questions, there are a number of issues which need to be clarified.

General consideration of analytical issues**Field test as sample.**

The sample for the remaining three analyses is the field test sample. This is appropriate since the field test sample represents a convenience sample (sixteen school districts in South Carolina which volunteered to be part of the field test, involving selecting and testing children who would fit the profile as explained below) of the specific population of interest, namely those children who fell into the pool eligible for the dimension identification protocol, and who scored highly enough in either Dimension A or Dimension B for identification, but remained unidentified because they did not score highly enough on both Dimension A and Dimension B.

Bias in general.

In shifting focus to consider questions of bias—the second, third and fourth of the research questions—it is appropriate to briefly consider what the empirical indicators of bias are. Some theoreticians would regard no difference between mean levels of performance as a prerequisite for lack of bias (Reynolds & Kaiser, 1990). Others would go even further:

Regardless of the purpose of a test or its validity for that purpose, a test should result in distributions that are statistically equivalent across the groups tested in order for it to be considered nondiscriminatory for those groups. (Alley & Foster, 1978; cited in Reynolds & Kaiser, 1990, p. 490)

After considering the positions advocated by such researchers, Reynolds and Kaiser (1990) declare that

“The mean difference and equivalent distribution concepts of test bias have been the most uniformly rejected of all criteria of test bias examined by sophisticated

psychometricians involved in investigating the problems of bias in assessment” (p. 490).

The implication of this firm declaration is that simplistic analyses of these Project STAR results are to be eschewed because the real issues raised by the specter of test bias is “the accuracy of ...labels across some nominal grouping system (typically race, sex, or socioeconomic status have been the variables of interest)”(Reynolds & Kaiser, 1990, p. 492; parentheses in original). The anchoring process, so central to the preceding question of reliability, was again required for the field test data to enable the results on Form A to be pooled with results on Form B at each of the three grade levels for which PACT data were available. The factor analytical approach preferred in these remaining three questions seeks to explore the factors that “statistically explain the variation and covariation among measures” (Green, Salkind, & Akey, 2000, p, 292) using exploratory factor analysis on a randomly selected dichotomy of the sample, sorted on the variable of interest, and to validate these factors by means of confirmatory factor analysis on the half of the sample not selected for the exploratory phase. By operating on the two randomly selected halves of the sample independently, this study avoids spurious validation which would arise if the exploratory analysis was then confirmed on the identical data.

In their further explication of the concept of bias, Reynolds and Kaiser (1990) sum up their position thus:

It is a question of whether race, sex, or any other demographic variable of interest influences the diagnostic process or the placement of a child in special programs independent of the child’s cognitive, emotional, and behavioral status. (Reynolds & Kaiser, 1990, p. 492)

General analytical criterion.

Hence the statistical expectation of an unbiased test in this and the remaining two questions addressed in this study will be that the factors nominated in the exploratory analysis with the randomly selected “first” half of the sample, sorted with respect to the three different variables of interest, will be validated in the confirmatory analysis performed on the “second” half. If identical factors are confirmed across the variables of interest, the Project STAR instrument will be evidencing a lack of bias with respect to that variable. While the issue of bias remains the focus, as the investigation of the data proceeds a number of other issues, particularly concerning the identification of factors, will emerge and be dealt with at those specific junctures.

Anchoring the field test sample

The field test sample consisted of 1786 children, of whom there were 1115 for whom “dense” data was available for Project STAR, in addition to the PACT data necessary for anchoring. Each distribution was graphed as a precaution against outliers, but, in contrast to the pilot data described earlier, there were no outliers sufficiently outstanding as to pose a threat to the anchoring process. The breakdown into categories of the 1115 children is shown in Table 8.

Table 8

Breakdown of field test sample, showing numbers of students for whom PACT data was obtained

Level	Grade	Form A	Form B
Intermediate	6	164	151
	5	181	196
Primary	4	216	207

It should be noted that these data apply to children nominally one grade older than the children in the pilot sample. This was done quite deliberately, since the field test was conducted early in the fall semester, when (it was reasoned) the children would be more closely allied to their preceding grade level than to their rising grade level in terms of maturity. Each of the three different grade levels (4, 5, and 6) must be anchored separately (see above), and this trichotomy of the sample will continue to be operative in the remaining analyses. After the anchoring process, exploratory factor analyses were performed on the combined Form B and Form B* (Form B equivalent for those who took Form A) data. An assumption underlying the use of combined results is that the two Forms were in fact parallel. While this issue has been addressed extensively in the previous section, it will be returned to later in this study.

The statistics required in the linear transform from Form A scores to Form B* scores are shown in Table 9.

Table 9

Descriptive statistics for equating Form A results to Form B equivalents- field test

Grade	Group	Statistic	Variable pertaining to:		
			Form A	Form B	PACT
4	A ₄	M _{4A}	24.96		627.83
		S _{4A}	5.21		19.02
		b _{4APACTA}	.186		
	B ₄	M _{4B}		29.82	631.02
		S _{4B}		4.14	14.57
		b _{4BPACTB}		.128	
	Total ₄	M ₄			629.39
		S ₄			17.04
5	A ₅	M _{5A}	35.07		830.28
		S _{5A}	6.68		16.24
		b _{5APACTA}	.219		
	B ₅	M _{5B}		37.13	832.34
		S _{5B}		6.17	16.53
		b _{5BPACTB}		.229	
	Total ₅	M ₅			831.35
		S ₅			16.40
6	A ₆	M _{6A}	38.26		1031.52
		S _{6A}	5.68		17.23
		b _{6APACTA}	.179		
	B ₆	M _{6B}		41.87	1033.61
		S _{6B}		6.69	16.59
		b _{6BPACTB}		.190	
	Total ₆	M ₆			1032.52
		S ₆			16.93

In Table 9, as in the preceding anchoring process, the subscript numbers refer to the grade level, and the subscript capital letters refer to the two forms: A and B. The “M” represents the mean and “S” the standard deviation. The “b” represents the slope of the regression line of the subscripted group on the PACT data for that group. Hence “ $b_{4APACTA}$ ” refers to the slope of the regression line of the fourth grade STAR results on the fourth grade PACT outcomes for those who took Form A of the STAR instrument.

The formulae for devising the variables involved in the linear transforms from Form A scores to Form B-equivalent (B*) scores using the statistics provided in Table 9 were cited above, and will not be repeated here. The equations for the linear transforms, at the respective grade levels were:

$$\text{For grade 4: } B^* = 0.86 (A - 25.25) + 29.61$$

$$\text{For grade 5: } B^* = 0.92 (A - 35.30) + 36.90$$

$$\text{For grade 6: } B^* = 1.19 (A - 38.44) + 41.66$$

Again, these formulae were used to generate the Form B-equivalent (B*) scores for those children who took Form A of the Project STAR instrument. The true and equivalent scores were combined for each grade, and independent t-tests were used to determine whether the two forms of the instrument were in fact yielding parallel results. With an alpha level of .05 and a two-tailed test at each grade level, the B* scores for those in grade 4 who took Form A of the Project STAR instrument ($M = 29.10$, $SD = 4.48$) were not significantly different from those who took Form B ($M = 29.81$, $SD = 4.14$), $t(421) = -1.70$, $p > .05$. Those in grade 5 who took Form A of the Project STAR instrument ($M = 36.68$, $SD = 6.14$) were not significantly different from those who took Form B ($M = 37.13$, $SD = 6.17$), $t(375) = -0.701$, $p > .05$. Similarly, those in grade 6 who took Form A of the Project STAR instrument ($M = 41.45$, $SD = 6.75$) were not

significantly different from those who took Form B ($M = 41.87$, $SD = 6.69$), $t(151) = -0.561$, $p > .05$. These results are summarized in Table 10.

Table 10

Summary of results for test of difference between B-equivalent (B*) and B scores after anchoring

Grade	B* mean (standard. deviation)	B mean (standard deviation)	t (df)	prob. (p)
4	29.10 (4.48)	29.81 (4.14)	-1.70 (421)	> 0.05
5	36.68 (6.14)	37.13 (6.17)	-0.701 (375)	>0.05
6	41.45 (6.75)	41.87 (6.69)	-0.561 (151)	>0.05

These results signify that with the field test sample, the two Forms A and B were in fact parallel forms and the differences in the scores on the tests were more reflections of the differences between the two samples than differences between the tests themselves. On the basis of this finding, the data will be analyzed without discriminating on the basis of the form of the test taken.

General exploratory factor analysis procedure

A random sample of approximately 50% of the total sample was selected to form the basis of the exploratory factor analyses. The designation of the cases in the exploratory factor analytic sub-sample was permanent, i.e. the cases identified as involved in the exploratory factor analyses were not included in the ensuing confirmatory factor analysis phase.

A number of different factor extraction methods were tested in accord with Johnson and Wichern's (1982) recommendation that "it is always prudent to try more than one method of solution. If the factor model is appropriate for the problem at hand,

the solutions should be consistent with one another" (p. 408). With these data, the principal component analysis method generally yielded a higher proportion of total variance explained by the factors with eigenvalues > 1 in the unrotated solution, and so was used uniformly at the factor extraction stage for all three grade levels. To address Johnson and Wichern's (1982) issue of consonance, it was noted that the other methods did show consistency with the principal component analysis. For example, at the grade 4 level, for eigenvalues > 1 , the number of factors and percentage of variance accounted for by those factors (in parentheses) were as follows: principal component analysis, 3 factors (52.08%); unweighted least squares, 3 factors (32.62%); generalized least squares, 3 factors (34.02%); maximum likelihood, 3 factors (32.7%); principal axis factoring, 3 factors (32.50%), and alpha factoring, 3 factors (32.57%). The consistency of these results support the use of the factor analytic technique with these data, while highlighting the effectiveness of principal component analysis at this extraction stage with these data.

Green, Salkind and Akey (2000) address the issue of how many factors to retain in factor analysis by recommending reflection throughout both the extraction (Stage 1) and rotation (Stage 2) phases on "(1) a priori conceptual beliefs about the number of factors based on past research or theory, (2) the absolute values of the eigenvalues computed in Stage 1, (3) the relative values of the eigenvalues computed in Stage 1, and (4) the relative interpretability of rotated solutions computed in Stage 2" (p. 294, numbering in the original). Johnson and Wichern (1982) advised that "the number of common factors retained in the model is increased until a 'suitable proportion' of the total sample variance has been explained" (p. 411, quotes in original). They went on to declare that "the best approach is to retain few rather than many factors, assuming they provide a satisfactory interpretation of the data and yield a satisfactory fit to S (the sample

covariance matrix) or \mathbf{R} (the sample correlation matrix)” (Johnson & Wichern, 1982, p. 411, parentheses added).

In this case, the scree test was consulted at the extraction stage in each of the three grades, with the number of factors corresponding to “the sharp descent part of the plot before the eigenvalues start to level off” (Green, Salkind & Akey, 2000, p. 297) being initially retained. Green, Salkind and Akey (2000) assert that “this criterion more frequently yields accurate results than the eigenvalues-greater-than-1 criterion” (p. 297). The scree plot for the grade 4 data is representative and is shown in Figure 7. It should be noted in passing that a two-factor solution agreed with the designation of the a priori designation of the two parts of the test as verbal and nonverbal at the grade 4 level, but was in contrast to the three designations of verbal, spatial and mathematical for the three parts of the test at the grades 5 and 6 levels. This issue will be returned to later. At this juncture, however, for each grade, two factors were rotated using a Varimax rotation procedure and the maximum likelihood methodology.

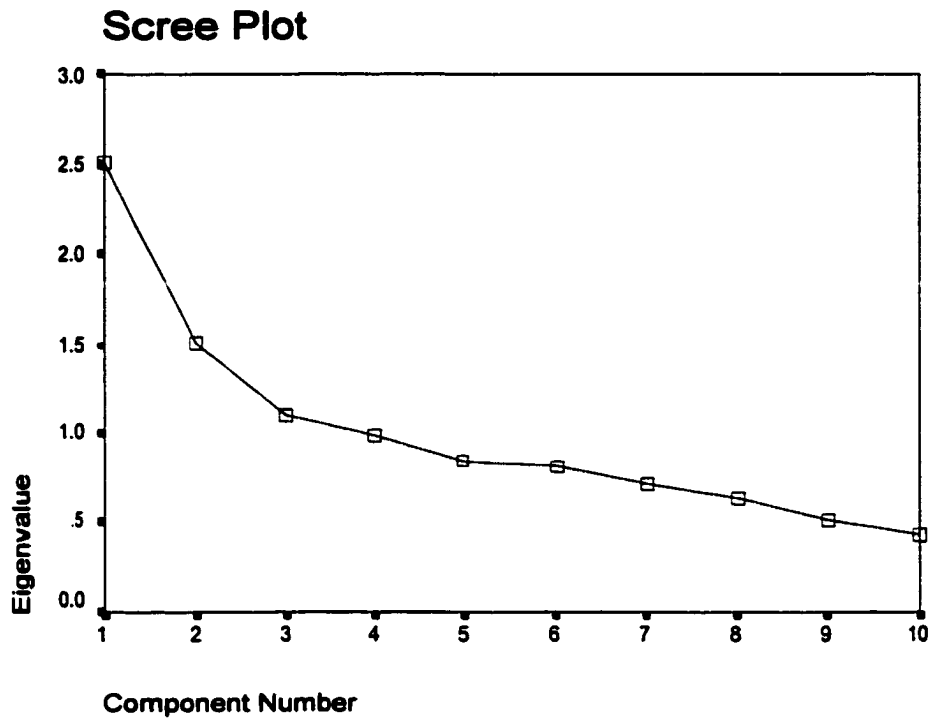


Figure 7 Scree plot for grade 4 exploratory data using principal component analysis: showing the two eigenvalues in the steep part of the graph.

For grade 4, the rotated solution yielded two interpretable factors named Factor 1 and Factor 2. Factor 1 accounted for 25.0% of the item variance, and Factor 2 accounted for 15.2%. One item was close to symmetrically bipolar. The loadings of the individual grade 4 items are shown in Table 11.

Table 11

Two-factor solution: Factor 1 and Factor 2 for grade 4

Item identifier	Factor 1	Factor 2
PV1	.63	-.17
PV2	.48	.34
PV3	.56	-.09
PV4	.52	.08
PV6	.55	.39
PNM1	.55	-.11
PNM3	.03	.67
PNS1	-.03	.68
PNS5	.62	.13
<i>Complexly Determined Item</i>		
PNS4 (<i>bipolar</i>)	.55	-.52

In Table 11 and the other summary tables, “P” signified a “Primary” item, “V” indicated a “Verbal” domain assignment, “N” a “Nonverbal” domain assignment, “M” a “Mathematical” domain assignment, and “S” a “Spatial” domain assignment. As Johnson and Wichern (1982) point out, “ideally we should like to see a pattern of loadings such that each variable loads highly on a single factor and has small-to-moderate loadings on the remaining factors. It is not always possible to get this simple structure...” (p. 423). While the factor loadings in Table 11 are satisfactory values from the fact that they meet the accepted criteria of being greater than .4, the reality is that only just over 40% of the variance in the data is accounted for by these two factors. As noted above, increasing the number of factors to 3 raised the percentage of variance accounted for to just over 50%, but this was achieved at the cost of interpretability—a cost which was deemed

unjustifiable in the face of a gain of only 10% to a figure which was still quite low. In view of the factor analysis outcomes for grade 4 and 5, it is noteworthy in passing that the item designation for the Primary level test originally encompassed separate mathematical and spatial domains, but that the two were combined with the eradication of excess items to form a nonverbal domain prior to the pilot phase. The item numbering preserved the original assignment of item to domain.

Two-factor solutions for the grade 5 and grade 6 data were generated similarly to that of grade 4, but in each case the initial 2-factor solutions were supplemented by the 3-factor solutions as shown below. In both of these grades, while the scree plot indicated a two-factor solution, a three-factor solution was found to be both more in agreement with the structure of the test and more readily interpretable. The three-factor solution for grade 5 (Table 12) follows.

Table 12

Three-factor solution: Factor 1, Factor 2, and Factor 3 for grade 5

Item identifier	Factor 1	Factor 2	Factor 3
IV3	-.01	-.06	.63
IV4	.47	-.03	.37
IV6	-.00	-.12	.43
IM1	.13	.65	-.21
IM7	.06	.76	.20
IM8	.40	.01	.10
IS4	.42	.22	.00
IS1	.45	.27	.07
IS3	.53	-.02	-.15
IS8	.42	.32	-.15

In Table 12, “I” signified an “Intermediate” item, “V” a “Verbal” domain assignment, “M” a mathematical domain assignment, and “S” a “Spatial” domain assignment. The combination of Factor 1, Factor 2, and Factor 3 accounted for only 25.6% of the variance in the grade 5 sample—a very low percentage. The 2-factor solution was worse, accounting for just under 20 % of the variance, as well as running counter to the a priori categorization which assigned each item to one of three domains. This issue is raised here as a prelude to a decision made at the confirmatory factor analytic phase.

The three-factor solution for grade 6 (Table 13) follows.

Table 13

Three-factor solution: Factor 1, Factor 2, and Factor 3 for grade 6

Item identifier	Factor 1	Factor 2	Factor 3
IV1	.00	.02	.25
IV4	.22	.04	.74
IM1	.71	-.14	.02
IM2	.05	.43	.18
IM7	.53	-.01	-.00
IM8	.30	.60	-.07
IS4	.54	.09	.02
IS1	.61	.27	.05
IS3	.33	.18	.03
IS2	-.00	.36	-.22
IS8	.58	-.01	.23

These three factors accounted for only 26.6% of the variance, and analogously to the comment on grade 5 above, the 2-factor solution accounted for just over 21% of the variance, in addition to being contrary to the a priori assignment of items to domains. As mentioned above, this facet of the analysis will be re-visited in the confirmatory stage.

Summary of exploratory factor analyses

To summarize the exploratory factor analytic results to date, the data from half of the total sample (randomly selected) at grade 4 level have supported the designation of two factors, and at the grades 5 and 6 levels have supported the designation of three factors. While the number of factors at each of these three levels is in agreement with the a priori determination of the number of domains, the discernible factors have run counter to the a priori designation of item membership of the factors. To accentuate this point, the items were designated in each of Tables 11, 12, and 13 in the order of presentation in the test booklet and with the original item identifiers attached. For grade 4 (Table 11), Factor 2 may be a verbal factor, but it included only two of the five items initially designated as verbal items, whereas Factor 1 would be more difficult to label. For grade 5 (Table 12), Factor 3 may have been verbal, and Factor 2 may have been mathematical, but with the same reduction of item membership. Again, Factor 1 would be more difficult to label. For grade 6 (Table 13), Factor 3 may have been verbal with reduced item membership, but both Factors 1 and 2 would be more difficult to label.

Confirmatory factor analysis

Finally this study has reached the confirmatory factor analysis phase in which the three key questions, the second, third and fourth of the research questions for the whole study, will be addressed. Specifically, it is crucial to this study to answer the question of bias in relation to gender, ethnicity, and socioeconomic status. If a dichotomy of the second half of the data on the basis of gender, for example, should show that the factorial structure that describes the responses of males is different from that which describes the response of females, then clearly the test has different attributes for the two genders, and this is clear evidence of bias. The same process and conclusion would pertain in the case

of the ethnicity dichotomy, and finally for the socioeconomic dichotomy.

In order to follow through this conceptual schema, the second half of the data not already utilized in the exploratory phase was used to try to account for the variance exhibited by the data at each of the grade 4, grade 5 and grade 6 levels. Having established the factor analytic characteristics of the Project STAR instruments at each level, the general approach was to take the appropriate subsets of each level in terms of gender, ethnicity and socioeconomic status and to try to model the variance in each sample upon the basis of the factors delineated at the exploratory phase. Should the Project STAR data exhibit a different structure for one subset as opposed to its complement, then there will be reason to suggest that the test is biased with respect to the dichotomy set up on that subset.

To implement this design, the data files residing in SPSS format were modeled in Amos, structural equation modeling software which stands alone as well as acting as a plug-in to the SPSS program itself. "Amos implements the general approach to data analysis known as structural modeling, analysis of covariance structures, or causal modeling" (Arbuckle, 1997, p. 1). There is no imperative to take the outcome of the exploratory factor analyses as input at this confirmatory phase. Consequently, the approach taken here was to analyze both the structure underlying the design of the instrument and the factor structure suggested by the exploratory factor analysis. In each of the following subsections the structure corresponding to the design assignment of items to domains will be given first, followed by the structure arising from the exploratory factor analysis reported above.

Grade 4

Figure 8 shows the standardized output of the structural equation model for the grade 4 Project STAR test instrument as the items were assigned to domains in the process of test construction, that is, on the basis of a priori judgment by competent judges. In Figure 8 the rectangles represent observed variables, and the ellipses represent unobserved variables.

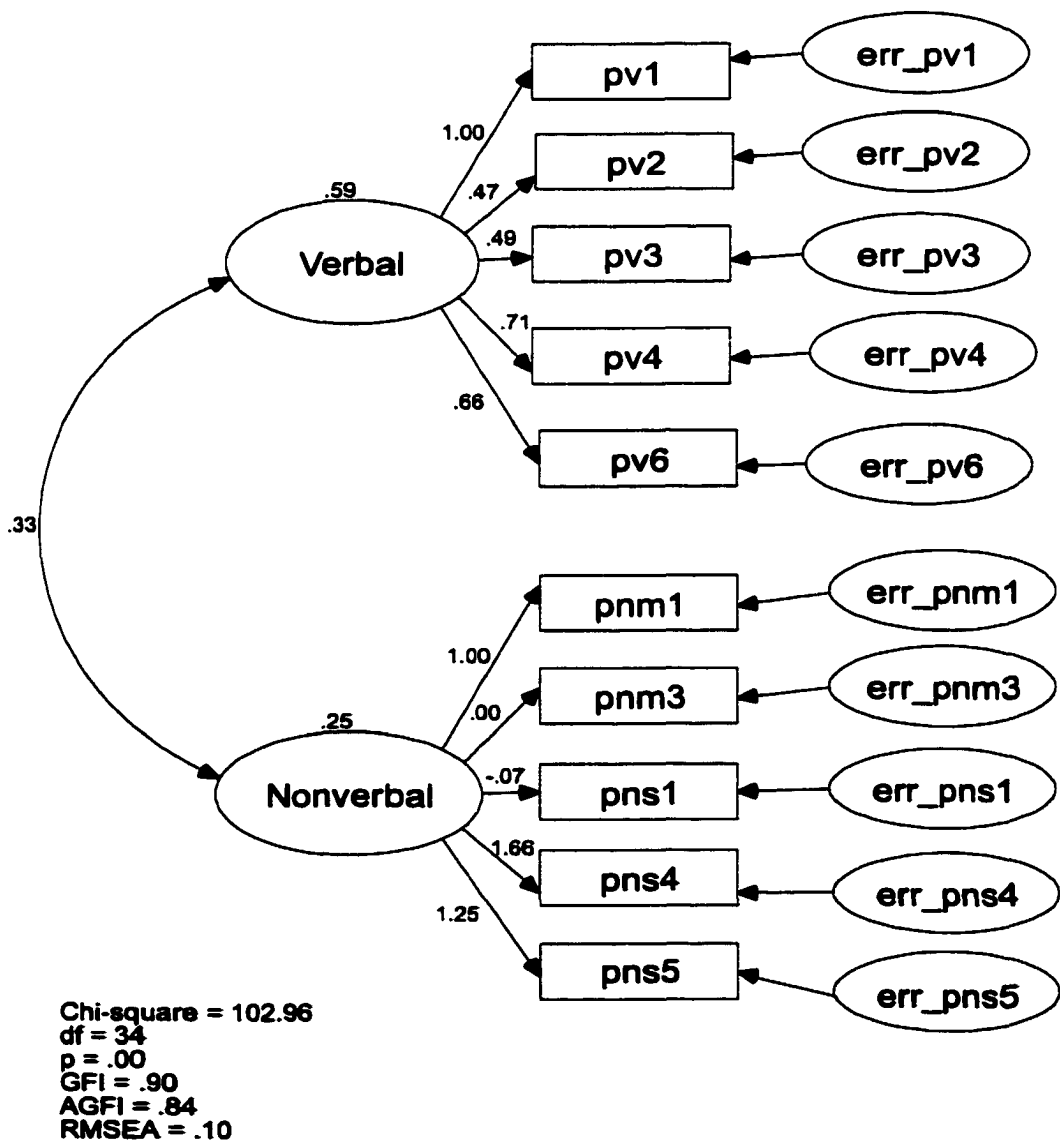


Figure 8 Structural equation model for a priori assignment of items to domains in the design process for the grade 4 Project STAR instrument.

The unobserved variables on the right hand side were labeled as error terms. This structure suggested that the score on the “pv1” item, for example, was partly attributable to a component emanating from a “verbal ability” construct, and partly to an error component, by which was meant, in this model, anything other than the “verbal ability” construct. The correlations between the “verbal ability” construct and the items appear next to the arrows in Figure 8. The covariance between the “verbal” and “nonverbal” constructs appear next to the double-headed arrow on the left hand side of the model.

The extent to which the data supported this model was measured by the chi-square value and by the goodness of fit (GFI), adjusted goodness of fit, (AGFI), and the root mean square error of approximation (RMSEA) indices. The chi-square value was a measure of the agreement between the implied and sample covariances where

the implied covariances are the best estimates of the population variances and covariances under the null hypothesis that the parameters required to have equal estimates are truly equal in the population, whereas the sample covariances are the best estimates obtained without making any equality assumptions. (Arbuckle, 1997, p. 328, italics in original)

The indicated value of chi-square, as shown in Figure 8 was 102.96 (df = 34). This was not likely ($p = 0.00$) if the null hypothesis was true. The remaining three indices were consistent with this judgment. The GFI index is always between zero and unity, where unity denotes a perfect fit. Here, the value for GFI was 0.90. The AGFI index takes into account the degrees of freedom available for testing the model and is bounded by 1 above, though not limited to 0 below as is the GFI. Here the AGFI was .84. Finally, to lend some perspective on this whole issue of goodness of fit, the RMSEA index adjusts for model complexity. Arbuckle (1997) cites Browne and Cudeck (1993) as

follows:

Practical experience has made us feel that a value of RMSEA of about .05 or less would indicate a close fit of the model in relation to the degrees of freedom. This figure is based on subjective judgement... We are also of the opinion that a value of about .08 or less for the RMSEA would indicate a reasonable error of approximation and (we) would not want to employ a model with a RMSEA greater than 0.1. (Browne & Cudeck, 1993; as cited in Arbuckle, 1997, p. 559)

In this case the RMSEA value was .10. This was right on the upper limit of what Brown and Cudeck (1993) would regard as “employable”, but when taken in conjunction with the other indicators, the overall impression as regards this model, taking into account the chi-square value and the other three indicators, was that it was not a good fit to the data.

Figure 9 shows the model which arose out of the direct application of the exploratory factor analysis of the grade 4 data. The items which loaded above .4 on each factor as recorded in Table 11, were related to those respective factors, leaving aside the PNS4 item, which was signified as being complexly determined.

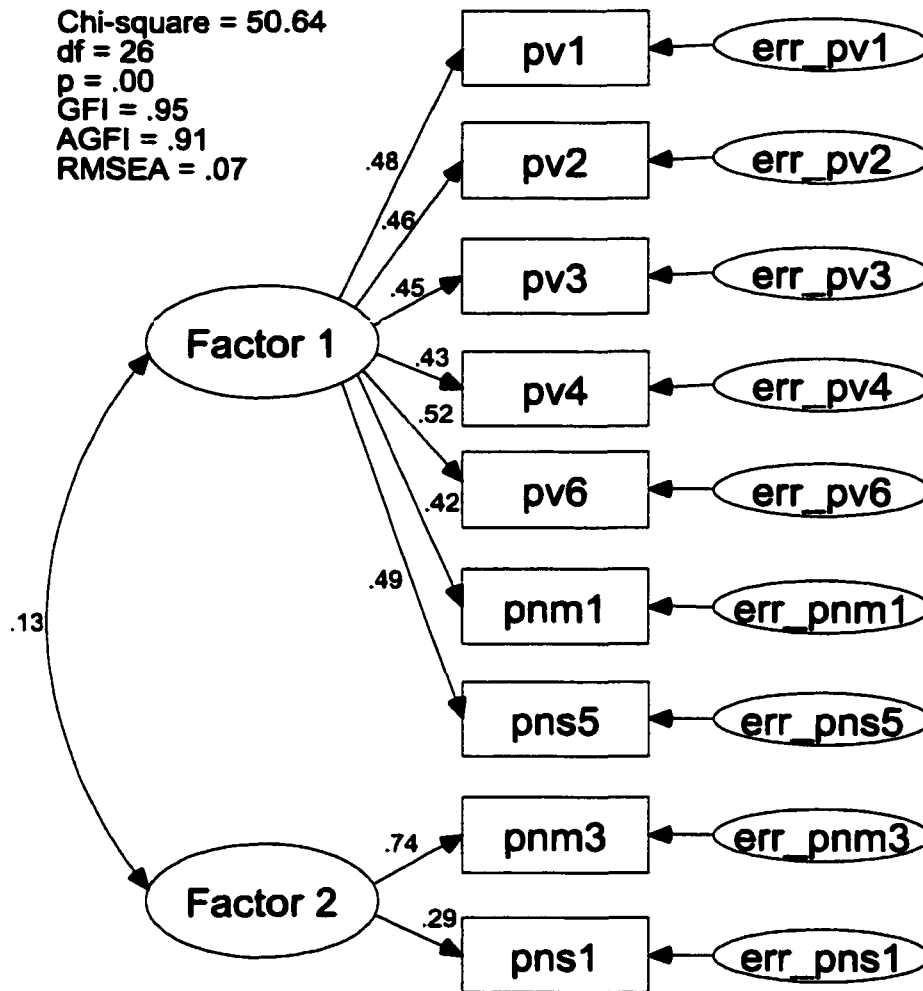


Figure 9 Two-factor solution for Grade 4, arising out of the exploratory factor analysis.

This was a better fit than the a priori model, as indicated by the GFI (which changed from 0.90 for the a priori model to 0.95 for this exploratory factor model) as well as the other indices, but it still failed to reach the level expected of a model showing good fit, as indicated by the low probability value.

Figure 10 showed the structural equation model when the exploratory factor analysis was taken as the basis for the model, as in figure 9, but after a process of elimination to find a model which was a good fit for the data. The second factor was

dropped, and the PNS5 item introduced, as with one factor only operating it is no longer complexly determined. For this model, the chi-square value of 8.44 (df = 9), $p = .49$ was quite likely if the null hypothesis was true. The GFI, at .99 was very close to 1, as was the AGFI, and the RMSEA at 0.00 agreed in indicating that this model was a good match for the data.

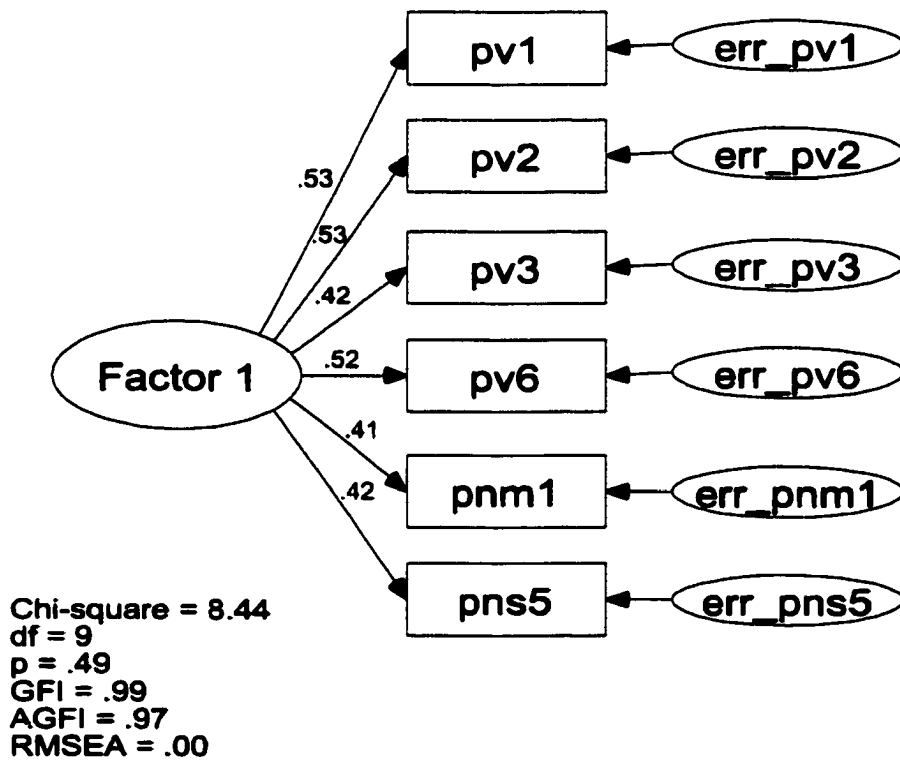


Figure 10 “Best-fit” structural equation model for grade 4 data based on exploratory factor analysis.

Grade 5

The situation with the a priori model for grade 5 was similar to that which pertained to the grade 4 data, except that the model did not allow the algorithm to reach convergence. This was an even better indication that the model did not fit the data. The input model is shown in Figure 11 to illustrate the structure of the model.

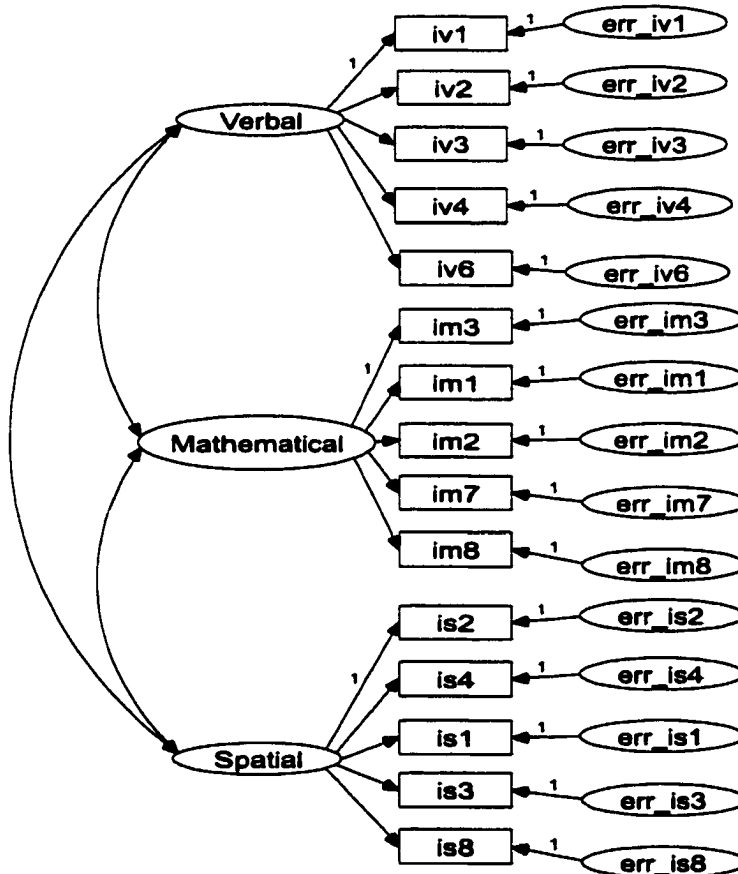


Figure 11 Structure of the priori model at grade 5 level.

In contrast, the 3-factor solution shown in Figure 12, predicated on the exploratory factor analysis, did converge, but the chi-square value of 47.04 ($df = 32$), $p < .03$ indicated that the null hypothesis was untenable. Also indicating that this model was not a good fit was the GFI of .94, the AGFI of .90, and the RMSEA of .06, a level for the RMSEA that does not reach Brown and Cudeck's (1993) criterion for a close fit model.

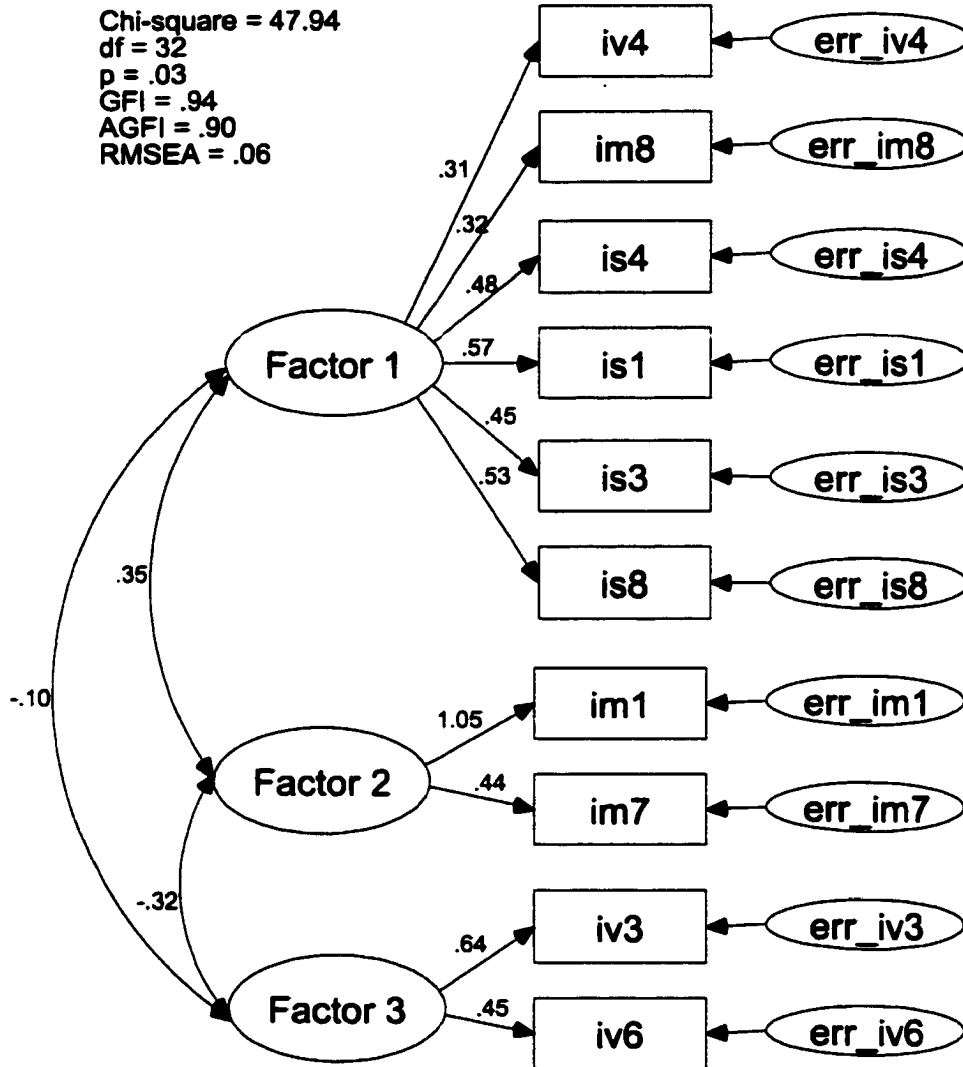


Figure 12 Three-factor model for the grade 5 data reflected the outcome of the exploratory factor analysis.

Because the scree test indicated only two factors, as mentioned above, the decision was made to analyze the grade 5 data on the basis of only two factors. The exploratory factor analysis for the 2-factor solution which was performed prior to deciding on the 3-factor solution was consulted. Interestingly, the 2-factor solution actually indicated a 1-factor solution because there was no item actually reaching a high enough loading on Factor 2 in this solution. In order to follow this train of investigation through, the data were fit to the model as shown in Figure 13.

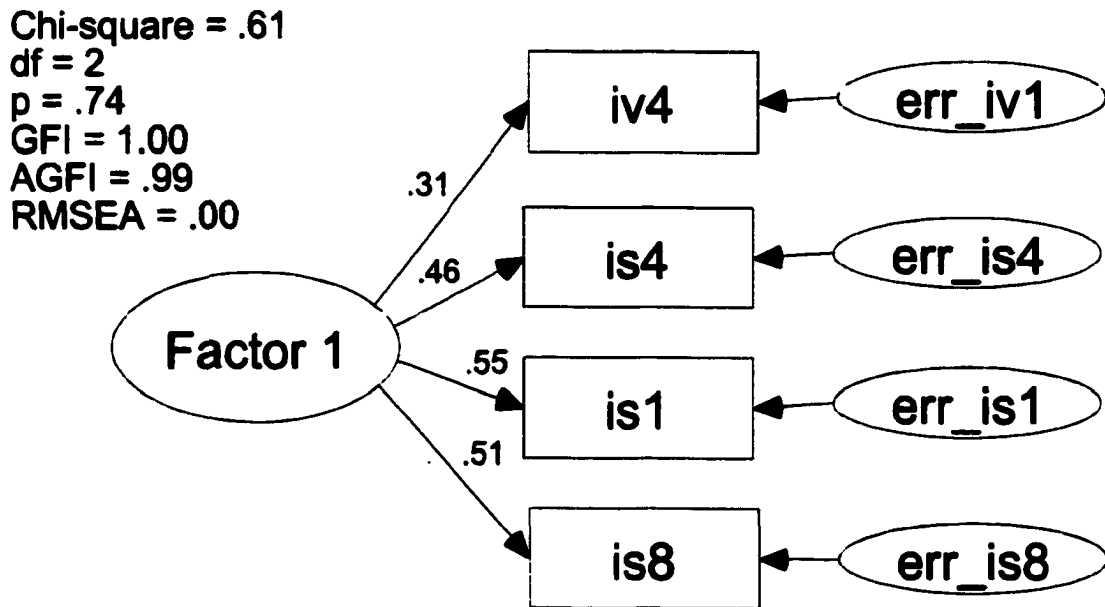


Figure 13 The 1-factor solution for the grade 5 data from the exploratory factor analysis.

In this somewhat reductionist version, the data again fit the model well, with the chi-square of 0.61 ($df = 2$), $p < .74$ in agreement with the GFI (1.00), the AGFI (.99) and the RMSEA (0.00).

Grade 6

The a priori grade 6 model converged, in contrast to the grade 5 model, but the chi-square of 153.88 ($df = 87$), $p = 0.00$ clearly indicated the lack of fit to the data, as did the GFI (.92), the AGFI (.89), and the RMSEA (.054). The exploratory factor analysis 3-factor solution became a 2-factor solution by default, because the third factor showed a loading on only one item, resulting in problems with the identification of the structure with respect to this third factor and its associated error term. This 2-factor default structure was not an acceptable model either with its chi-square of 102.51 ($df = 14$), $p = .000$ in consonance with with the GFI (.92), the AGFI (.84) and the RMSEA (.16).

Findings on bias

As Keith and Reynolds (1990) observed, “bias denotes constant or systematic error, as opposed to chance or random error in the estimation of some value; in test bias research, this constant or systematic error is usually the result of group membership or some other nominal variable...” (p. 52). The above results have shown no evidence of any constant or systematic error. Keith and Reynolds (1990) went on to attest that “bias exists in regard to construct validity of a test whenever that test can be shown to measure different hypothetical traits or constructs for one group than it does for another group, or to assess the same construct but with differing degrees of accuracy” (p. 52). In this regard, again the Project STAR instrument was lacking bias when it portrayed a similar factor structure for each of the sub-sample dichotomies on the variables of interest: namely, gender, ethnicity and socioeconomic status. The factor structure of the Project STAR instrument was not clear-cut, and this has reflected been reflected in the indeterminate outcomes which emerged in some instances.

Table 14 presented a summary of the outcomes of the analyses of the bias investigations, in terms of the statistics related to the best fit model.

Table 14

Model fit statistics bearing on the issue of bias

Sample subset	Chi-square	df	p	GFI	AGFI	RMSEA
g4gmale	6.81	9	.66	.969	.929	.000
g4gfemale	13.28	9	.15	.963	.913	.065
g5gmale	3.00	2	.23	.975	.876	.091
g5gfemale	3.08	2	.21	.982	.908	.082
g6gmale	52.02	14	.00	.910	.821	.149
g6gfemale	63.71	14	.00	.907	.814	.162
g4ewhite	6.90	9	.65	.984	.962	.000
g4eaa	12.94	9	.17	.916	.805	.100
g5ewhite	0.611	2	.74	.998	.989	.000
g5eaa	0.751	2	.69	.983	.917	.000
g6ewhite	97.90	14	.00	.903	.806	.173
g6eaa	67.32	19	.00	.803	.710	.228
g4syes	9.23	9	.42	.904	.861	.025
g4sno	9.63	9	.38	.977	.947	.023
g5syes	0.32	2	.85	.995	.975	.000
g5sno	0.26	2	.88	.999	.994	.000
g6syes	64.84	19	.00	.805	.712	.215
g6sno	80.24	14	.00	.914	.829	.156

The interpretation of the sample subset column was as follows: the first two characters indicated the grade level, the next one character indicated whether the subset was on the basis of gender (g), ethnicity (e), or socioeconomic status (s), and the remaining characters indicated which dichotomy was analyzed on the male/female, white/african-american, and free/reduced lunch status of the child (on a yes/no basis).

In contrast with the lengthy preparatory phases, the outcome of the final stage of

this study, wherein the crucial questions of this study were addressed, is starkly simple. Despite the issues discussed earlier concerning the difficulty establishing factor structure, when it came to the final phase, the structural equation model fit the grade 4 and grade 5 data. Only the grade 6 model failed to fit the data, and it failed regardless of which of the dichotomies it was applied to. This outcome still validates a claim of non-bias, albeit in a negative way: the best available model fit neither gender, neither ethnicity, and neither socioeconomic status group.

Concluding remarks

While the meaning of the various results has been explained as they were reported above, it is helpful at the close of this chapter to recap on what has been found. Firstly, it was shown that the Project STAR instruments were psychometrically appropriate for this implementation. To do this first necessitated the use of a third test as an anchor test to relate the outcomes on Form A of the test with the outcomes on Form B of the test. Following the appropriate re-grouping of the results, high Cronbach alpha values were typically found, but not values which would be typical of commercially-developed psychometric instruments. It was concluded in terms of some extenuating theoretical conditions and the projected use of Project STAR, that the instruments were appropriate. Statistics were reviewed which validated the intention that Project STAR should both be somewhat similar to “usual” tests employed to gauge student progress, and yet quite distinct from these same tests. The psychometric question was pursued on data from a heterogeneous sample, namely the pilot test group.

A homogeneous sample (i.e. the field test group) distinct from the pilot group was used for all further analyses. The aim of this second phase of the study was to address the general issue of bias by analyzing the factor structure of the Project STAR

instruments with various sub-samples of interest. Before this could be done, linear transforms arising out of anchoring the Project STAR outcomes on the PACT data validated the conflation of the testing results at all levels. Then a random selection of half the sample responses was factor analyzed at each of the three grade levels (grade 4, 5 and 6). The pervasive difficulty concerned the lack of any particularly satisfying factor model which could account for more than half of the variance in any of the grade-level samples.

Finally, the a priori factors which underpinned the design of the test, and the factors from the exploratory phase were inserted into a set of structural equation models with mixed, but generally satisfactory results. For grades 4 and 5, the structural model fit all sub-samples well; for grade 6 the structural model fit none of the sub-samples well. Regardless of this outcome, there was no evidence of bias in the field test sample of the Project STAR instruments.

The discussion, conclusions, and implications for the identification of high-ability learners of what has been discovered in the analysis of these data is the concern of the ensuing Chapter 5 discussions. In pursuing this, Chapter 5 will also return to the underpinning themes developed in the literature review.

Chapter Five

Conclusions

Introduction

This study set out to investigate a large sample of carefully scored data from a well-controlled administration of a new type of group-administered, performance-based assessment instrument developed as part of Project STAR. In this instrument children were taught a relatively novel task and then asked to perform on a very much similar task. Children who were able to do this were dubbed high-ability learners. The theory supporting this technique to detect high-ability learners was detailed as arising from the work of Vygotsky (1978). His concept of the zone of proximal development theoretically underpinned the expectation that at least some children would be able to perform at the new advanced level of understanding because their zones of proximal development were sufficiently expansive to encompass this performance when mediated by a teacher. While not part of this study, this expectation was fulfilled. Students were spread across a continuum by the use of these instruments, with many satisfying the arbitrary identification criterion set in the South Carolina regulations, of four out of five items correct.

The focus of this study was not the numbers who were identified, but the characteristics of the test itself. This focus gave rise to the research questions concerning the establishment of the psychometrics of the two forms of the test in this context. Establishing the basic psychometrics would validate its use in addressing three key questions which had to do with the issue of bias.

The question of the basic psychometrics of the test necessitated addressing the question of whether the results on the two Forms of the test were equivalent. Results on the administration of the test to a large heterogeneous sample (pilot administration) were used to conduct an anchoring process to determine whether the results on the two forms

could be regarded as parallel. This determined that for this psychometric phase, Form A and Form B should be considered separately at the primary level, but were parallel at the intermediate level. Out of these investigations, the psychometrics of the instrument were reported as satisfactory. All of this constituted the prelude to the three main questions: Do the outcomes on the Project STAR instruments exhibit a bias on the basis of gender, ethnicity, or socioeconomic status?

The main body of this study concerned the field test sample and issue of bias. The results for a sample of children selected on the basis of satisfying the requirements on either Dimension A or Dimension B of the South Carolina Gifted and Talented Regulation 43-220 (amended May 12, 1999) , but not both, were subjected to the anchoring process utilized in addressing the psychometrics. This indicated that the results on Form A and Form B were equivalent, and hence the outcomes at each of the three grade levels involved were combined. The sample was then randomly dichotomized, and exploratory factor analysis was conducted on one of the halves. The factors discerned in this exploratory factor analysis phase were then used as the basis for confirmatory factor analysis by way of structural equation modeling in the ensuing phase with the second of the halves of the sample.

Discussion

Reynolds and Kaiser (1990) proposed that bias in testing has been “a recurring social embroglio throughout the history of mental measurement” (p. 487). They were clearly referring to bias along ethnic lines, as they went on to refer to “emotionally laden polemics decrying the use of mental tests with any minority group member who has not been exposed to the cultural and environmental circumstances of the white middle class” (p. 487). Some of the most encouraging aspects of this study, as noted in the previous

chapter, arose at the denouement of the analysis when the less-than-robust models that were nevertheless the best that could be devised on the total grade level samples fit very well to the data when they were dichotomized along the three dimensions that most readily come to the fore when the specter of bias is raised: gender, ethnicity, and socioeconomic status.

Ultimately this study was successful in showing that there was no evidence of bias in the field test implementation of Project STAR. The same structural equation model outcomes held for each of the dichotomies along the lines of gender, ethnicity, and socioeconomic status at each of the three grade levels.

Contribution of this study

The bar was set high at the start of this study when what was being sought was nominated as being evidence that there is some propensity being tapped by the Project STAR instruments which could be identified as high-ability learning. This has not been achieved. If this had been found, it would have been an exceptionally strong recommendation for the use of the Project STAR instruments. Empirical evidence from sub-human physiological psychology has been reviewed to validate the concept of high-ability learning in that field, but the hoped-for statistical indication of a similar ability at work in the responses to the Project STAR instruments has not been forthcoming.

It was hoped that Project STAR would constitute a defensible identification procedure, correlated with other instruments of achievement and ability, and yet distinct from them. This was achieved. The Project STAR instruments are worthy of endorsement in the context of their use.

The desire to see Project STAR established as a part of the identification process in South Carolina has come to pass, due mainly to the determination of key individuals in

South Carolina to press on with the development of the instruments, rather than to obsess over shortcomings. In this regard, the major contribution of Project STAR to date has been the addition of a non-biased test to the identification menu. It has been clearly established that the same factors operate for the children from low socioeconomic status backgrounds that operate for the children not from such backgrounds. In itself, this is a major contribution, although too much enthusiasm is tempered by the fact that the factor structure is not robust, and that at grade 6 level the best factor structure fails to fit the data for either of the socioeconomic dichotomies.

It was hoped that this study would yield valuable insights into the potential of performance-based assessment to place pens into the hands of those well suited to wield them. A dispassionate assessment would have to acknowledge that such a specific claim could not be made at present. Certainly Project STAR is identifying children who are able to perform well on its items. But the fine-grained detail that was hoped for has proved elusive.

Intangible contributions from Project STAR have been far-reaching. Many teachers have become involved in looking again at the identification decisions that have been made. The act of reviewing past decisions has brought to light some incorrect judgments, and these have been corrected. Many teachers have become engaged in administering the Project STAR tests, and this has involved them very directly with items that stretch their understanding, and which require them to use a sound pedagogical method, and to come to terms at some level with sound theory for educating high-ability learners. The long-term value of such ownership should not be discounted.

Remaining issues

The first question that deserves to be addressed is why there was such difficulty in

delineating factors in a sample comfortably large enough to sustain such an analysis, and from data which were collected using a test instrument which was consciously designed with a factorial structure at its very heart. This is a complex question which defies a simple answer, but it is possible to make some suggestions, while at the same time noting that, although it is a finding to be concerned about, the fact that there was no “obvious” factor structure in a particular instance of a test is not in itself a condemnation of that test. That being said, there are two aspects of this study worthy of consideration in this context.

It is possible that the items simply didn’t relate to the domains, or that, in this context, the items were acting in a manner contrary to the way they would customarily be perceived in a testing situation. A more plausible explanation was that the anchoring process, while totally defensible from a statistical point of view, had led to a conflation of factors. Another explanation was simply that this implementation was anomalous in its outcome. In that case, subsequent implementations will be far more easily interpreted. The small proportion of variance accounted for by the identified factors would be expected to make some at least of the confirmatory analyses problematic. This issue has been pervasive enough to be referred to as the “elusive factor” issue.

It was decided to explore further the suggestion that the anchoring process may have been a factor in contributing to the difficulty in delineating factors. The intent here is not to second-guess any of the discussion so far, but to appropriately pursue the viability of a plausible cause of the elusive factor issue. Firstly, to recap, the decision to anchor the Form A and Form B results by using the PACT data as the anchor was made on the advice of a noted theoretician who knew the South Carolina population, and both the PACT and Project STAR test well. It was significant that from among a number of

others at the table at the time that decision was made the objection immediately arose that the PACT was not a good instrument for anchoring purposes for a number of reasons. One reason had to do with the PACT being in its infancy itself. Another of the reasons related to the philosophical issue that, being an achievement test itself, PACT could not be expected to correlate highly with Project STAR. This objection gave weight to the result of analyses between MAT-7 and Project STAR showing mostly significant positive correlations of approximately .3 between the MAT-7 testing result and the Project STAR outcomes. The suggestion at that time had been to use either the MAT-7 or the OLSAT for anchoring as both had a longer established record as tests, and results on both were available. Nevertheless, it was decided to use the PACT results, and, as noted above, the PACT results on the field test showed that the two Forms of Project STAR were parallel forms.

A further suggestion at that time, which was not adopted, was to include some common items in the two forms which could form the basis of the anchoring process. One of the reasons against this was the fact that the tests were felt to be too long as they stood, although they consisted of only five items in each of two or three domains for primary and intermediate levels of the test respectively. South Carolina wanted two distinct forms of the test and it was felt that the common items if actually part of the test, would start to blur the distinctive nature of the forms.

The question that deserved to be addressed by this closing analysis related to whether the anchoring may have confounded the outcome by validating the combining of scores which were comparable in terms of some proportion of their inherent variance, but were also distinct in terms of a large proportion of their variance. This was advanced as a plausible explanation for the small proportion of the variance accounted for by the factors

in the exploratory factor analysis stage. The reasoning was that once this small proportion of the variance was all that could be extracted from the data at the exploratory stage, that large proportion of unexplained variability ensured difficulty in validating a factorial structure at the confirmatory stage. The most straightforward way to proceed in relation to this objection was to analyze the forms of the Project STAR test separately.

Consequently, the grade 4, 5, and 6 exploratory sample data were separately processed using exploratory factor analysis. The results did not confirm the hypothesis, with the grade 4 two-factor solution accounting for 26% of the variance on Form A and 23% on Form B, and the three factor solutions with grades 5 and 6 accounting for approximately 30% on both Form A and Form B.

This invalidated what had appeared to be the most plausible argument for the “elusive factor” issue. As already mentioned, Project STAR is currently in the planning stage for an imminent state-wide implementation. It will be extremely interesting to see if the factor structure is more clear as a result of the improvements made from the field test to the state-wide implementation, scheduled to take place between May 8 and May 19, 2000 on Form A only.

While the lack evidence for a clear factorial structure is vexing, a second aspect of this study which has bearing on the “elusive factor” issue has to do with the inherently complex nature of performance-based assessment—a point to which this concluding chapter will return. The verbal and non-verbal labels do not identify simple constructs in this context. Fundamentally, in this particular implementation of the performance-based learning paradigm, a child is required to learn a new skill in a single exposure and to then demonstrate a deep understanding of the underlying concepts involved in the skill, relying not on established learning, but on recently established connections. Children who

were successful with these performance-based items have certainly demonstrated an ability which could be aptly described as high-ability learning. Along this line of thought, if one abandons the scree plot as a guide to the delineation of the number of factors, one finds that a single factor solution including six of the ten items at the grade 4 level produces a chi-square of 8.44 ($df = 9$) with an associated probability level of .49 ($GFI = .99$, $AGFI = .97$, & $RMSEA = .00$). However, a single factor does not underlie either the remaining four of the ten items at grade 4 level, nor the entire ten items of the test. This single-factor pattern holds at the grade 5 level for six out of the fifteen items, and at the grade 6 level for five out of the fifteen items. This single factor may be thought of as a general factor, but in this context, it may also be thought of as a “high-ability learning” factor. If so, why does this factor not apply to all the items, or at least to all the items in the domain? Here we are brought back to the issue raised already in connection with the inability to detect factors accounting for the “usual” proportion of variance.

A final alternative explanation for the inability to delineate factors needs to be raised, and that possibly the items on the test instrument simply did not relate to the domains to which they were assigned. Here one must be careful not to overgeneralize from a single set of test results, bearing in mind the complexity of the response modality required of children on this test. Certainly, on the face of it, the items did relate to the domains as assigned. The items were reviewed many times by a number of different experienced and well-credentialed subject-matter experts who evinced no concern about the domain assignment of items. Moreover, items were deliberately designed for specific domains, and even specific prototypes within these domains. Perhaps it is that the dynamic of performance-based assessment in this group setting introduces different nuances from those which operate in the traditional classroom setting, resulting in

different aspects of items becoming predominant. The whole issue of factor structure remains a vexing, but open question.

While they should not overshadow the preceding findings, a number of intriguing issues arose in the course of this study and its discussion. The commitment of the South Carolina State Department of Education to the ongoing development of Project STAR will assure a fresh set of data in the near future. It will obviously be preferable to pursue “elusive factors” with a new set of data if for no other reason than it should not be thought that any one particular instance of a test administration is representative of all administrations—a point which has been made a number of times in the course of this study. The availability of fresh data also defuses any impulse to reiterative processing of the existing data until a more plausible solution emerges. One of the major benefits of the next wave of data is that it will all be from the one form of the test (Form A).

A very powerful recommendation for an educational test is its track record; its long-term reputation as a test which detects some characteristic highly valued in educational circles in the culture. The term “high-ability” when qualifying “learning” designates just such a highly valued characteristic. Clearly a test in its infancy cannot reference a track record, but if the Project STAR instrument remains in use, such data will become available if a follow-up component is added.

Moreover, teachers should be given resources and training in how best to stimulate the learning of a student identified using the Project STAR instrument. It would seem obvious that if a child is identified using Project STAR, but had difficulty with one or the other of Dimension A or B, some modification of the usual program for the gifted and talented is indicated. South Carolina is well aware of this reality and is taking steps to empower teachers to modify the curriculum to allow such children to achieve at a high

level.

The question was raised when the underpinning concepts were being discussed in the coda of the literature review as to whether the concept of a high-ability learner should be regarded as a hypothetical construct--something which would be referred to in factor analytic terms as a latent variable--or whether it in fact could be regarded as a correlate of an identifiable physiological advantage in terms of structure. In the literature review, extensive discussion was entered into concerning the evidence that high-ability learning is indeed more than a hypothetical construct. Somewhat disappointingly, this study has not produced evidence for a single underlying factor among the set of performance-based instruments at any of the grade 4, 5, or 6 levels. One might well expect evidence to arise for a general underlying factor ("g" factor), especially if the Project STAR test were functioning as a test of ability. In one sense, then, it is reassuring that there is no such factor, since Project STAR lays no claim to being a test of ability per se. And yet if the analogy from sub-human to human species outlined above is to hold fast, one would expect there to be a common factor underlying all items. Continuing effort would be expected to shed more light on the critical question of exactly how the Project STAR instruments operate, and in linking this understanding to the exciting research into the physiological substrate of high-ability learning.

The concept of dramatism

Because this study has failed to show the action of a persistent factor across each of grades 4, 5 and 6, the concept of high-ability learning has not been statistically identified. The concept of dramatism, however, remains as a powerful underpinning concept for the methodology which has been depicted as a modified dynamic assessment protocol. Dramatism grew out of Burke's "attempt to avoid the limitations, and even the

arrogance, of...monist perspectives” (Wertsch, 1998, p. 13). The social milieu which gave rise to Project STAR and the interplay of motivations enmeshed in the identification issue can be encompassed by using the lens of dramatism to avoid simplistic interpretation of the human complexity inherent in the Project STAR undertaking. For example, one dramatic scenario could be played out as follows.

Act: A child is taught how to perform a task by a more skilled adult, who then requests the child to perform a similar task.

Scene: The usual classroom, with a teacher who is keen to do a good job, but is also motivated by a sense of the injustice that may be perpetrated if this child is not identified because he/she is sure that the child deserves to receive extra help.

Agent: The child, who is by no means keen to be identified as “one of them.”

Agency: The Project STAR manipulative materials and response sheet.

Purpose: Ostensibly, the purpose is to identify someone worthy of special services, but the situation is conflicted. The teacher believes that the child *should* try harder. The child wishes that the teacher wouldn’t push so hard. The school principal, though an indirect player, has as agenda, of which the teacher is aware, because he/she may not have a class for such children, or a teacher who is both competent and keen to teach high-ability learners, or may apprehend political fall-out if one child is identified and another is not.

Of course, very different scenarios could be developed. It would not be unreasonable to suggest that each case would represent a different dramatic scenario. The beauty of the lens of dramatism is the kaleidoscopic perspective it lends to the analysis of action.

At the heart of the learning involved in Project STAR lies the concept of mediated

action. The child in the Project STAR test has a more highly skilled adult to point the way, to mediate learning for the child by pointing out where value inheres. Wertsch (1998) develops his own theory of mediated action, out of the foundation of dramatism, as his own personal perspective on sociocultural analysis. The point of the sociocultural approach, Wertsch (1998) says is “to explicate the relationships between human *action*, on the one hand, and the cultural, institutional, and historical contexts in which this action occurs, on the other” (p. 24, italics in the original). This approach is also a powerful lens in this context.

As intimated at various stages of this study, the Project STAR initiative arose as a very pointed response out of a cultural, institutional, and historical context which is uniquely South Carolinian. At both a global and a local level there are stakeholders. The Office of Civil Rights needs to be convinced that efforts to redress a proportional imbalance along racial lines are in hand. Meanwhile, the South Carolina legislature does little to reassure the doubters by wrestling with conflict over the flying of the Confederate flag from the state capitol dome. The people just this year returning from New Jersey to the Clarendon 1 school district after having their house burned down in the unrest surrounding the Brown vs Board of Education decision (Evans, D., 2000, personal communication) want to know if it is still “business as usual,” or whether the situation has changed. It would indeed be a serious error to misjudge the importance of who wields the cultural tools.

But Wertsch (1998) goes further. He wishes to stress that there is a downside to mediated action. The learner is restricted to learning what the teacher is teaching. Vygotsky viewed “the development of language in human ontogenesis primarily in terms of how it provides new capacities for human consciousness” (Wertsch, 1998, p. 38).

Wertsch seeks to emphasize that “if a new cultural tool frees us from some earlier limitation of perspective, it introduces new ones of its own” (p. 39). To Burke, such limitations may be terministic screens: “culture and language not only open doors to experiences, they also form a prison which constricts and narrows” (Gusfield, 1989, p. 12). The essence of being a teacher is to be an essentially benign guide to the learner, to be aware, at both a very general and a very elemental level, of the fact that “there are no negatives in nature, and that this ingenious addition to the universe is solely a product of human symbol systems” (Burke, 1966, p. 9).

Conclusion

This study showed that it is possible to identify young children from among those who test inconsistently on usual measures of ability and achievement who can perform in a way which would indicate that they may well have great potential for learning. It did not show that the potential for learning manifested by these children could be identified as “high-ability learning,” or attributable to any other single construct. It did show that group-administered, performance-based assessment, as defined and implemented in this test, evidenced no bias along the lines of gender, ethnicity, or socioeconomic status. In the course of showing the above, this study also showed that this group-administered, performance-based test exhibited satisfactory psychometric characteristics, and possessed concurrent validity to the degree expected with usual measures of ability and achievement. Taken as a whole, this is an encomium for the use of performance-based measures as part of an identification protocol for selecting those to be eligible for services as gifted learners.

Implications

The overriding implication of this study is that performance-based identification

should be utilized as part of the testing battery available to school districts seeking to assess potential for learning. While the high-ability learning construct has not been demonstrated statistically, the physiological evidence that such a construct exists at least in sub-human species, and the reality of the distribution of a quite select group of students on the basis of the outcome of carefully-designed performance-based instruments combine to validate the use of such an instrument at least as a modulator of standard ability and achievement tests. The lack of bias, particularly in terms of ethnicity and socioeconomic status, is a further strong recommendation for using performance-based assessment as an alternative approach.

For the practitioner, the power of the teaching methodology cannot go unnoticed. This study relied on the power of instruction along the lines of the “zone of proximal development.” Very large numbers of children not deemed “gifted” by the yardstick of ability-and-achievement proved themselves quite capable of handling demanding learning exercises with aplomb. It is worth mentioning again the observation of a number of teachers involved in Project STAR that children became engrossed in the performance tasks and did not want to move on to the next item.

For the researcher, this study illustrated some of the difficulties that arise in testing complex concepts. The existence of a genetically manipulable physiological process analogous to high-ability learning at the sub-human level, did not mean that it was statistically identifiable with a human sample. The fact that the performance-based items were so carefully designed did not mean that even that structure was detectable in the final outcome.

Project STAR illustrated the need to persist in following a developmental path in operationalizing complex concepts, specifically because the outcome of complex

processes may defy simple explanations. When complex processes are involved, perhaps it is advisable to simplify the administration by using a single form, although this comment is offered only in this context, and would not be being made if the factor structure had been less elusive.

Such were the vistas that were opened by the endeavor that is Project STAR. They are grand, but they are no grander than the vistas that are opened by the cumulative effect of every act of teaching. The potential for learning embedded in every human being is latent until it is evoked by a skilled teacher. For some, it would seem that there is greater than usual potential. The greater the potential, the greater the obligation to detect it, to evoke it, and to nurture it to fruition—to indeed place the pen in the hand well qualified to wield it.

Appendix A

**Sample items in the Project STAR format
(culled in the design and selection process)**

Domain: Verbal Running 2V5B
Prototype: Concept Development
Level: Primary Grades 2 & 3
Materials: Copies of Picture
Preteaching Example:

Show students the following picture:



Ask them to suggest a title for the picture in the space provided, and describe the situation on the lines provided. Ask for a few responses. Check to be sure that everyone understood the task by asking if there are any questions about what they were asked to do. Be sure students understand to write a narrative response rather than a list.

Tell students to turn to "Running 2V5B." Read the directions aloud and tell them to begin.

Preteaching example



Title:

Description:

Running 2V5B

Name: _____

Create a funny title for the following picture and describe why you think it is funny.



Title:

Description:

Rubric for Task 2V5B (Pictorial/verbal humor) Kicking

4	3	2	1	0
Both title and paragraph reflect strong understanding of pictorial humor.	Both title and picture reflect good understanding of pictorial humor.	Title is humorous but paragraph is limited in being able to explain humor.	Both title and paragraph lack understanding of pictorial humor.	No response

Note to scorers:

Multiple answers prevail. You may wish to sort a set of student papers into two piles (strong vs. weak) and then sort into four piles in order to apply the rubric effectively.

Students may write an analytical explanation of their title or a humorous story. Either approach should receive full credit.

Domain: Math Hamburgers 5M5A
Prototype: Proportional Reasoning
Level: Intermediate Grades 4 & 5
Materials: (None)

Preteaching example:

1. Say, "You are going to work a problem that involves the idea of ratio which is a kind of matching of groups. Here is an example: Matti is putting a bouquet of flowers together. The ratio of flowers to leaf stalks in the bouquet is 5 to 2. This means that for every 5 flowers she uses there are 2 stalks of leaves. It doesn't mean that the only possibility is that she has 5 flowers and 2 leaves. She could have 10 flowers and 4 leaves, or any other number pair where every 5 flowers matches to 2 leaf stalks.

Write this table on the board. Say, "Here are some possibilities."

Flowers	Leaf Stalks
5	2
10	4
15	6

"Look at the third possibility. This means if we make groups of 5 flowers, there are 3 groups of 5 with none left over. We can match 2 leaf stalks to each group of flowers making 6 leaf stalks and then there are no leaf stalks left over."

2. Have students turn to "Hamburgers 5M5A." Read the problem with the students and tell them to begin.

Hamburgers 5M5A

Name: _____

Some students from Eagle School are at a picnic. There are 83 hamburgers. This is enough for each student to have at least one. For every 3 boys there are 2 girls.

1. How many students could be at the picnic? How many are boys? How many are girls?

(List all possible answers. Show or write how you got the answers.)

2. What is the largest possible number of students who could be at the picnic? How many are boys ? How many are girls ?

3. What is the smallest number who could be at the picnic ?

Rubric for 5M5A (Proportional reasoning) 83 hamburgers

4	3	2	1	0
At least 8 points; largest and smallest numbers identified.	5-7 points; largest and smallest numbers may or may not be identified.	2-4 points.	1-2 points.	No response.

Notes to scorers:

Give one point for each correct pair given (see table below).

The largest number of boys is 48 and girls is 32 .

The smallest number of boys is 3 boys and girls is 2.

Students	Boys	Girls
80	48	32
75	45	30
70	42	28
65	39	26
60	36	24
55	33	22
50	30	20
45	27	18
40	24	16
35	21	14
30	18	12
25	15	10
20	12	8
15	9	6
10	6	4
5	3	2

Spatial: Grades 2-3
Prototype: Patterning
Materials: Square tiles, handout

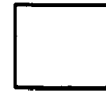
Practice Item: Give out tiles. Do the following practice pattern with students. Place one tile on overhead projector. Say, "this is the first part of a pattern."



Place 3 more tiles on overhead in the following pattern:



Say, "This is the second part of a pattern. How many tiles did we add? Where did we put them?"



Place 3 more tiles on the overhead in the following pattern:



Say, "This is the third part of a pattern. How many tiles did we add? Where did we put them?"

"Now can you use your tiles to make the next part of the pattern?" (Allow students time to arrange tiles into next part of the pattern.)

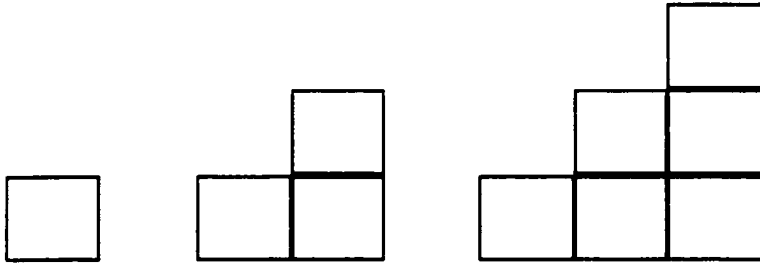


"This is the pattern you should have formed."

"Now you are going to try another pattern." Pass out handout and read the directions with students.

Name: _____

Study the following pattern. Draw the next two patterns. You may use your tiles to help you.



Scoring for Patterning

- 5** correctly draws 4 and 5; responds correctly to number of tiles for each (4 is 10, 5 is 15)
- 4** correctly draws 4 and 5; responds with correct number in one of two blanks
- 3** 4 and 5 correct in drawing but tile numbers are omitted in blank
- 2** 4 or 5 is correct in drawing ; # of tiles is 10 for # 4 or 15 for #5, but incorrectly filled in blank
- 1** no response, both 4 and 5 are incorrect

References

- Alamprese, J.A., Erlanger, W.J., & Brigham, N. (1988). No gifts wasted: Effective strategies for educating highly able disadvantaged students in mathematics and science. Washington, DC: United States Department of Education.
- Anastasi, A. (1992). Review of the Otis-Lennon School Ability Test, Sixth Edition. In J.J. Kramer, & J.C. Conoley (Eds.), The Eleventh Mental Measurements Yearbook (pp. 633-635). Lincoln, NE: The University of Nebraska.
- Arbuckle, J.L. (1997). Amos users' guide, version 3.6. Chicago: SmallWaters Corp.
- Austen, J. (1996; originally published 1818). Persuasion. New York: Book of the Month Club.
- Baldwin, A.Y. (1985). Programs for the gifted and talented: Issues concerning minority populations. In F.D. Horowitz & M. O'Brien (Eds.), The gifted and talented: Developmental perspectives (pp. 223-249). Washington, DC: American Psychological Association.
- Baldwin, A.Y. (1991). Ethnic and cultural issues. In N. Colangelo & G.A. Davis (Eds.), Handbook of gifted education (pp. 416-427). Boston: Allyn & Bacon.
- Baldwin, J.M. (1906). Social and ethical interpretations in mental development: A study in social psychology. New York: Macmillan.
- Barnett, W.S. (1998). Long-term effects on cognitive development and school success. In W.S. Barnett & S.S. Boocock (Eds.), Early care and education for children in poverty: Promises, programs, and long-term results (pp. 1-39). Albany: State University of New York Press.
- Borland, J.H. (1989). Planning and implementing programs for the gifted. New

York: Teachers College Press.

Borland, J.H., Schnur, R., & Wright, L. (2000). Economically disadvantaged students in a school for the academically gifted: A postpositivist inquiry into individual and family adjustment. Gifted Child Quarterly, 44(1), 13-32.

Borland, J.H., & Wright, L. (1994). Identifying young, potentially gifted, economically disadvantaged students. Gifted Child Quarterly, 38(4), 164-171.

Brown, S.W., & Yakimowski, M.E. (1987). Intelligence scores of gifted students on the WISC-R. Gifted Child Quarterly, 31, 130-134.

Burke, K. (1966). Language as symbolic action: Essays on life, literature, and method. Berkeley: University of California Press.

Burke, K. (1968). Dramatism. In D.L. Sills (Ed.), International encyclopedia of the social sciences, (vol. 7, pp. 445-447). New York: Crowell Collier and Macmillan,

Burke, K. (1969a). A grammar of motives. Berkeley: University of California Press.

Burke, K. (1969b). A rhetoric of motives. Berkeley: University of California Press.

Burke, K. (1972). Dramatism and development. Worcester, MA: Clark University Press.

Burke, K. (1984). Attitudes toward history. Berkeley: University of California Press.

Campione, J.C., Brown, A.L., Ferrara, R.A., & Bryant, N.R. (1984). The zone of proximal development: Implications for individual differences and learning. In B. Rogoff & J.V. Wertsch (Eds.), Children's learning in the "zone of proximal development" (pp. 77-92). San Francisco: Josey-Bass.

Carroll, J.B. (1997). Psychometrics, intelligence, and public perception.

Intelligence, 24(1), 25-52.

Charlotte-Mecklenberg Public Schools. (1993). Magnet program student enrolment and participation: Report for Charlotte-Mecklenberg Schools. Charlotte, NC: Author.

Chittenden, E. (1991). Authentic assessment, evaluation, and documentation of student performance. In V. Peronne (Ed.). Expanding student assessment, (pp. 22-31). Alexandria, VA: Association for Supervision and Curriculum Development.

Cisek, G.J. (1998). Filling in the blanks: Putting standardized tests to the test. Fordham Report, 2(11), 1-51.

College Board, The. (1999). Reaching the top: A report of the National Task Force on Minority High Achievement. New York: College Board Publications.

Crocker, L., & Algina, J. (1984). Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.

Csikszentmihalyi, M. (1991). Flow: The psychology of optimal experience. New York: HarperCollins.

Csikszentmihalyi, M., Rathunde, K., & Whalen, S. (1993). Talented teenagers: The roots of success & failure. Cambridge, UK: Cambridge University Press.

Darby, D. (2000). Member of South Carolina State Steering Committee for Project STAR. Private communication.

Darling-Hammond, L. (1995). Equity issues in performance-based assessment. In M.T. Nettles, & A.L. Nettles (Eds.). Equity and excellence in educational testing and assessment, (pp. 89- 114). Boston: Kluwer Academic.

Educational Policies Commission. (1950). Education of the gifted. Washington,

DC: National Education Association and American Association of School Administrators.

Eignor, D.R., & Hambleton, K.R. (1979). Effects of test length and advancement score on several criterion-referenced test reliability and validity indices. Amherst, MA: Laboratory of Psychometric and Evaluative Research Report No. 86, School of Education, University of Massachusetts.

Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). Rethinking innateness: A connectionist perspective on development. Cambridge, MA: Bradford Books with MIT Press.

Entwistle, D.R., & Hayduk, L.A. (1978). Too great expectations. Baltimore, MD: Johns Hopkins University Press.

Federal Interagency Forum on Child and Family Statistics. (1998). America's children: Key national indicators of well-being, 1998. Federal Interagency Forum on Child and Family Statistics, Washington, DC: U.S. Government Printing Office.

Feuerstein, R. (1979). The dynamic assessment of retarded performers. Baltimore, MD: University Park Press.

Feuerstein, R., Rand, Y., & Hoffman, M. (1979). The dynamic assessment of retarded individuals: The learning potential assessment device, theory, instruments, and techniques. Glenview, IL: Scott Foresman & Co.

Finley, C. J. (1995). Review of the Metropolitan Achievement Tests, seventh edition. In J.C. Conoley, & J.C. Impara (Eds.), The Twelfth Mental Measurements Yearbook (pp. 603-606). Lincoln, NE: The University of Nebraska.

Fishkin, A.S., Kampsnider, J.J., & Silverman, L.K. (1997). WISC-III patterns of cognitive abilities for gifted and high-ability children. (Unpublished paper delivered at

National Association for Gifted Children Convention, Little Rock, AR.)

Frasier, M.M. (1980). Programming for the culturally diverse. In J. Jordan, & J. Grossi (Eds.), An administrator's handbook on designing programs for the gifted and talented. Reston, VA: Council for Exceptional Children.

Frasier, M. (1987). The identification of gifted black students: Developing new perspectives. Journal for the Education of the Gifted, 10(3), 155-180.

Frasier, M.M. (1993). Issues, problems and programs in nurturing the disadvantaged and culturally different talented. In K.A. Heller, F.J. Monks, & A. H. Passow (Eds.), International handbook of giftedness and talent (pp. 685-692). Oxford: Pergamon.

Frierson, E. (1965). Upper and lower status gifted children: A study of differences. Exceptional Children, 32(2), 83-90.

Gagne, F. (1997). Critique of Morelock's (1996) definitions of giftedness and talent. Roeper Review, 20(2), 76-85.

Gall, M.D., Gorg, W.R., & Gall, J.P. (1996) Educational research: An introduction (6th ed.). White Plains, NY: Longman.

Gallagher, J.J. (1985). Teaching the gifted child. Newton, MA: Allyn and Bacon.

Gallagher, J.J. (1998, November). Unthinkable thoughts and gifted education: A commentary. Unpublished address delivered at National Association for Gifted Children Convention, Louisville, KY.

Gallagher, J.J., & Gallagher, S.A. (1994). Teaching the gifted child. Boston: Allyn and Bacon.

Glassman, M. (1994). All things being equal: The two roads of Piaget and Vygotsky. Developmental Review, 14, 186-214.

Gottfredson, L.S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. Intelligence, 24(1), 13-23.

Gottfried, A.W., Gottfried, A.E., Bathurst, K., & Guerin, D.W. (1994). Gifted IQ: Early developmental aspects. The Fullerton Longitudinal Study. New York: Plenum.

Green, S.B., Salkind, N.J., & Akey, T.M. (2000). Using SPSS for Windows: Analyzing and understanding data (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Gusfield, J.R. (Ed.). (1989). Kenneth Burke on symbols and society. Chicago: University of Chicago Press.

Hambleton, R.K. (1990). Criterion-referenced testing methods and practices. In T.B. Gutkin, & C.R. Reynolds (Eds.), The handbook of school psychology (pp. 388-415). New York: John Wiley & Sons.

Hambleton, R. K. (1995). Review of the Metropolitan Achievement Tests, seventh edition. In J.C. Conoley, & J.C. Impara (Eds.), The twelfth mental measurements yearbook (pp. 606-610). Lincoln, NE: The University of Nebraska.

Hansen, S.L., & Ginsburg, A. (1986). Gaining ground: Values and educational success among disadvantaged students. Washington, DC: United States Department of Education.

Hollinger, C.L., & Kosek, S. (1986). Beyond the use of full scale IQ scores. Gifted Child Quarterly, 30 (2), 74-77.

Howley, A., Howley, C.B., & Pendarvis, E.D. (1986). Teaching gifted children: Principles and strategies. Boston: Little, Brown and Company.

Hunsaker, S.L. (1994). Adjustments to traditional procedures for identifying underserved students: Successes and failures. Exceptional Children, 61(1), 72-76.

Ingels, S.J., & Quinn, P. (1996). Sample exclusion in NELS:88: Characteristics of

base year ineligible students: changes in eligibility status after four years. Washington, DC: United States Department of Education, Office of Educational Research and Improvement.

James, W. (1958). Talks to teachers. New York: Norton.

Jencks, C. (1972). Inequality. New York: Basic Books.

Jitendra, A.K., & Kameenui, E.J. (1993). An exploratory study of dynamic assessment involving two instructional strategies on experts and novice's performance in solving part-whole mathematical word problems. Diagnostique, 18, 305-324.

Johnson, B. (1996). The performance assessment handbook (Vols. 1-2). Princeton, NJ: Eye on Education.

Johnson, R.A., & Wichern, D.W. (1982). Applied multivariate statistical analysis. Englewood Cliffs, NJ: Prentice-Hall.

Kaufman, A.S. (1992). Evaluation of the WISC-III and the WPPSI-R for gifted children. Roeper Review, 14, 154-158.

Kay, S.I., & Subotnik, R.F. (1994). Talent beyond words: Unveiling spatial, expressive, kinesthetic, and musical talent in young children. Gifted Child Quarterly, 38, 70-74.

Kearney, K., & LeBlanc, J. (1993). Forgotten pioneers in the study of gifted African Americans. Roeper Review, 15, 192-199.

Kirschenbaum, R.J. (1998). Dynamic assessment and its use with underserved gifted and talented populations. Gifted Child Quarterly, 42(3), 140-147.

Kuhn, T. S. (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.

Lazar, I. (1981). Early intervention is effective. Educational Leadership, 38, 303-

305.

Lee, S.S., & Lord, E.W. (1999). Notification of Amended Regulations.

Unpublished memorandum to District Superintendents and Principals (6/1/99). Office of Curriculum and Standards, Department of Education, State of South Carolina.

Lidz, C.S. (1987). Historical Perspectives. In C.S. Lidz (Ed.), Dynamic assessment: An interactional approach to evaluating learning potential (pp. 3 - 32). New York: Guilford Press.

Lidz, C.S. (1991). Practitioner's guide to dynamic assessment. New York: Guilford.

Linn, R.L., & Niemi, D. (1995). Dimensionality and generalizability of domain-independent performance assessments. The Journal of Educational Research, 89(4), 197 - 205.

McIntosh, M.E., & Greenlaw, M.J. (1986). Fostering the postsecondary aspirations of gifted urban minority students. Roeper Review, 9(2), 104-107.

McLaughlin, M.W., & Shepard, L.A. (1995). Improving education through standards-based reform: A report by the National Academy of Education Panel on Standards-Based Educational Reform. Stanford, CA: The National Academy of Education.

Marcon, R.A. (1999, March). Impact of parent involvement on children's development and academic performance: A three-cohort study. Paper presented at the Meeting of the Southeastern Psychological Association, Savannah, GA.

Marland, S.P., Jr. (1971). Education of the gifted and talented. Volume I: Standards-based educational reform. Stanford, CA: The National Academy of Education.

Minick, N. (1987) Implications of Vygotsky's theories for dynamic assessment. In

C.S. Lidz (Ed.), Dynamic assessment: An interactional approach to evaluating learning potential (pp. 116-140). New York: Guilford

Natriello, G., McDill, E.L., & Pallas, A.M. (1990). Schooling disadvantaged children: Racing against catastrophe. New York: Teachers College Press.

Office of Educational Research and Improvement, United States Department of Education. (1993). National excellence: A case for developing America's talent. Washington, DC: U.S. Government Printing Office.

Passow, A.H. (1982). The gifted disadvantaged: Some reflections. In Identifying and educating the disadvantaged gifted and talented. Selected proceedings from the Fifth National Conference on Disadvantaged Gifted and Talented. Los Angeles: The National/State Leadership Training Institute on the Gifted and Talented.

Passow, A.H. (1986). Educational programs for minority/disadvantaged gifted students. In L. Kanevsky (Ed.), Issues in gifted education.. (pp. 147-172). San Diego, CA: San Diego City Schools GATE Program.

Passow, A.H. (1989). Needed research and development in educating high ability children. Roeper Review, 11(4), 223-229.

Passow, A.H., & Frasier, M.M. (1996). Toward improving identification of talent potential among minority and disadvantaged students. Roeper Review, 18(3), 198-202.

Patchett, R.F., & Stansfield, M. (1992). Subtest scatter on the WISC-R with children of superior intelligence. Psychology in the Schools, 29, 5-10.

Paulson, F.L., Paulson, P.R., & Meyer, C.A. (1991). What makes a portfolio a portfolio? Educational Leadership, 48(5), 60-63.

Pendarvis, E., & Howley, A. (1996). Playing fair: The possibilities of gifted education. Journal for the Education for the Gifted, 19(2), 215-233.

Piaget, J. (1970). Piaget's theory. In P.H. Mussen (Ed.), Carmichael's manual of child psychology (Vol.1, 3rd ed., pp. 703-732). New York: Wiley.

Plomin R., & Petrill, S.A. (1997). Genetics and intelligence: What's new? Intelligence, 24(1), 53-77.

Ramey, C.T., Yeates, K.O., & Short, E.J. (1984). The plasticity of intellectual development: Insights from preventive intervention. Child Development, 55, 1913-1925.

Raven, J.C. (1987). Standard Progressive Matrices. In D.J. Keyser, & R.C. Sweetland (Eds.), Test Critiques (Vol. 1, pp. 595-602). Kansas City, MI: Test Corporation of America.

Renzulli, J., Smith, L., & Reis, S. (1982). Curriculum compacting: An essential strategy for working with gifted students. Elementary School Journal, 82, 185-194.

Renzulli, J.S., Smith, L.H., White, A.J., Callahan, C.M., & Hartman, R.K. (1976). Scales For Rating the Behavioral Characteristics of Superior Students. Storrs, CT: Creative Learning Press.

Reynolds, C.R., & Kaiser, S.M. (1990). Test bias in psychological assessment. In T.B. Gutkin, & C.R. Reynolds (Eds.), The handbook of school psychology (pp. 487-525). New York: John Wiley & Sons.

Richert, E.S. (1987). Rampant problems and promising practices in the identification of disadvantaged gifted students. Gifted Child Quarterly, 31, 149-154.

Roedell, W.C., Jackson, N.E., & Robinson, H.B. (1980). Gifted young children. Perspectives on gifted and talented education. New York: Teachers College Press.

Rowe, D.C. (1997). A place at the policy table? Behavior genetics and estimates of family environmental effects on IQ. Intelligence, 41(1), 133-158.

Royce, J., Lazar, I. Darlington, R.B. (1983). Minority families, early education

and later life chances. American Journal of Orthopsychiatry, 53(4), 706-720.

Seitz, V., Rosenbaum, L.K., & Apfel, N.H. (1985). Effects of family support intervention: A ten year follow-up. Child Development, 56, 376-391.

Silver, S.J., & Clampitt, M.K. (1990). WISC-R profiles of high ability children: Interpretation of verbal-performance discrepancies. Gifted Child Quarterly, 34, 76-79.

Silverman, L.K. (1997). Issues in identification and assessment. (Unpublished paper presented at the 12th World Conference of the World Council for Gifted and Talented Children, Seattle, WA. July, 1997.)

Sternberg, R.J. (1993) Procedures for identifying intellectual potential in the gifted: A perspective on alternative "metaphors" of mind. In K.A. Heller, F.J. Mönks, & A.H. Passow (Eds.), International handbook of research and development of giftedness and talent (pp. 185-207). Tarrytown, NY: Pergamon.

Sternberg, R.J. (1998). Abilities are forms of developing expertise. Educational Researcher, 27(3), 11-20.

Sternberg, R.J., & Zhang, L. (1995). What do we mean by giftedness? A pentagonal implicit theory. Gifted Child Quarterly, 39(2), 88-94.

Swerdlik, M. E. (1992). Review of the Otis-Lennon School Ability Test, Sixth Edition. In J.J. Kramer, & J.C. Conoley (Eds.), The Eleventh Mental Measurements Yearbook (pp. 635-639). Lincoln, NE: The University of Nebraska.

Tannenbaum, A.J. (1983). Gifted children: Psychological and educational perspectives. New York: Macmillan.

Thelen, E., & Smith, L.B. (1994). A dynamic systems approach to the development of cognition and action. Cambridge, MA: Bradford Books with MIT Press.

Thorndike, E.L. (1914). The psychology of learning. New York: Teachers College

Press.

Tudge, J.R.H., & Winterhoff, P.A. (1993). Vygotsky, Piaget, and Bandura: Perspectives on the relations between the social world and cognitive development. Human Development, 36, 61-81.

United States Department of Agriculture. Food and Nutrition Service. (1999). Child nutrition programs—Income eligibility guidelines. Federal Register, 64(63), 1591-1592.

United States Department of Education. National Center for Educational Statistics. (1998). Parent involvement in children's education: Efforts by public elementary schools. Washington, DC: Author.

United States Department of Education. National Commission of Excellence in Education. (1983). A nation at risk: An imperative for educational reform. Washington, DC: Author.

United States Department of Health and Human Services. Office of the Assistant Secretary for Planning and Evaluation. (2000). The 2000 HHS poverty guidelines. Federal Register, 65(31), 7555-7557.

United States General Accounting Office. (1993). Student testing: Current expenditures, with cost estimates for a national examination. (Report No. GAO-PEMD-93-8). Washington, DC: U.S. Government Printing Office.

van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky, and beyond. Psychological Review, 105(4), 643-677.

VanTassel-Baska, J. (1989). The role of the family in the success of disadvantaged gifted learners. In J. VanTassel-Baska & P. Olszewski (Eds.), Patterns of influence: The home, the self, and the school (pp. 66-80). New York: Teachers College

Press.

VanTassel-Baska, J. (1991). Research on special populations of gifted learners. In M.C. Wang, M.C. Reynolds, & H.J. Walberg (Eds.), Handbook of special education: Research and practice. Vol. 4. Emerging programs (pp. 77-101). Oxford: Pergamon Press.

VanTassel-Baska, J., Patton, J., & Prillaman, D. (1989). A national survey of programs and practices for at risk learners. In Focus on Exceptional Children, (pp. 1-16). Denver, CO: Love.

VanTassel-Baska, J., & Willis, G. (1987). A three year study of the effects of low income on SAT scores among the academically able. Gifted Child Quarterly, 31, 169-173.

Venezky, R.L. (1992). Textbooks in school and society. In P.W. Jackson, (Ed.), Handbook of research on curriculum (pp. 436-464). New York: Macmillan.

Vinovskis, M. A. (1999). Do federal compensatory education programs really work? A brief historical analysis of Title 1 and Head Start. American Journal of Education, 107, 187-209.

Virginia Department of Education (1996). The Virginia plan for the gifted: A guide for the development of effective program services for gifted students. Richmond, VA: Author.

Vygotsky, L.S. (1978). Mind in society: The development of higher psychological process (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Trans., Eds.). Cambridge, MA: Harvard University.

Vygotsky, L.S. (1987). The collected works of L.S. Vygotsky: Vol. 1. Problems of general psychology (R.W. Rieber & A.S. Carton, Eds., N. Minick, Trans.). New York:

Plenum.

Wertsch, J.V. (1998). Mind as action. New York: Oxford University Press.

Wiggins, (1990). The case for authentic assessment. ERIC Digest. Washington, DC: ERIC Clearinghouse on Tests, Measurement, and Evaluation.

Wilkinson, S.C. (1993). WISC-R profiles of children with superior intellectual ability. Gifted Child Quarterly, 37, 84-91.

Wright, L., & Borland, J.H. (1993). Using early childhood developmental portfolios in the identification and education of young, economically disadvantaged, potentially gifted students. Roeper Review, 15, 205-210.

Zorman, R. (1997). Eureka: The cross-cultural model for identification of hidden talent through enrichment. Roeper Review, 20, 54-61.

Vita**Robert Martin Reardon**

- Birthdate:** January 05, 1949
- Birthplace:** Charters Towers, Queensland, Australia
- Education:** 1967-1969 Mt St Mary Teachers College
Sydney, New South Wales, Australia
Certificate of Teaching
- 1968-1977 Macquarie University
Sydney, New South Wales, Australia
Bachelor of Arts
- 1983-1984 Queensland University of Technology
Brisbane, Queensland, Australia
Graduate Diploma in Computer Education
- 1986-1990 University of Queensland
Brisbane, Queensland, Australia
Bachelor of Educational Studies
- 1991-1995 Queensland University of Technology
Brisbane, Queensland, Australia
Master of Education
- 1996-1997 University of New South Wales
Kensington, New South Wales, Australia
Certificate of Gifted Education

Abstract

ASSESSING POTENTIAL FOR LEARNING: A FACTOR-ANALYTIC STUDY OF A PERFORMANCE-BASED IDENTIFICATION PROTOCOL FOR YOUNG, SOCIOECONOMICALLY DISADVANTAGED HIGH-ABILITY LEARNERS

Reardon, Robert Martin, Ph. D. The College of William and Mary in Virginia, 2000. 173 pp.

Chair: Professor Joyce L. VanTassel-Baska

This factor-analytic study of a performance-based identification protocol for young, socioeconomically disadvantaged high-ability learners investigated the issues of reliability, test equivalency, and bias. A group-administered, performance-based set of instruments was designed in a joint project between the Center for Gifted Education and the State Department of Education, South Carolina. These instruments went through a series of processes of review and refinement leading to their use in a field test in fall 1999. The outcome of this field test administration is the subject of exploratory and confirmatory factor analysis in this study.

Reliability of the instruments was established on the pilot study data which were gathered from a heterogeneous sample of 1425 students. Statistical anchoring using linear transforms was used to address the status of the two forms of the test instruments. The Cronbach alpha values ranged from 0.71 to 0.78, values lower than desirable for psychometric instruments, but acceptable in view of the special purpose of this test. Exploratory factor analysis on a randomly chosen half of the field test data (N = 1800 students) lead to structural equation modeling of both a priori and exploratory factors on the second half of the field test data.

The exploratory factor analyses did not support a construct of high-ability learning. All emergent factors accounted for less than a majority of the variance in the relevant sub-samples. Nonetheless, the structural equation models demonstrated that there was no evidence of bias on the basis of gender, ethnicity, or socioeconomic status. Project STAR did indeed exhibit the ability to discriminate in an unbiased way among young, socioeconomically disadvantaged high-ability learners.

The overriding implication of this study is that performance-based identification should be utilized as part of the testing battery available to school districts seeking to assess potential for learning. At the same time, the failure to detect a strong factorial structure in the results of a performance-based test specifically designed around a factorial schema implies that there are layers of complexity inherent in this testing protocol that deserve close attention. Further research arising from increasingly standardized implementations is expected to shed more light on what has been called in this study the "elusive factor" issue.