A COMPARISON BETWEEN THE DRAW A PERSON AND THE STANFORD-BINET INTELLIGENCE TEST IN PREDICTING SCHOOL ACHIEVEMENT

By

DEBRA F. MARCOTTE

Bachelor of Arts

The University of Tulsa

Tulsa, Oklahoma

1992

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
July, 1994

OKLAHOMA STATE UNIVERSITY

A COMPARISON BETWEEN THE DRAW A PERSON AND THE STANFORD-BINET INTELLIGENCE TEST IN PREDICTING SCHOOL ACHIEVEMENT

Thesis Approved:

Ruth E. James
Thesis Adviser

Laura Hubbs-Tait

Mona Lane

Thomas C. Collins
Dean of the Graduate College

ACKNOWLEDGMENTS

I would like to thank Dr. Ruth Tomes, my adviser and mentor, for her support and guidance throughout my graduate program. Dr. Tomes offered encouragement and went out of her way to help me with my masters thesis. The support and suggestions of Dr. Mona Lane and the statistical advice of Dr. Laura Hubbs-Tait were invaluable. I would also like to thank Dr. Ruth Ann Erdner and her colleagues for all of their help during the data collection process.

Without the support and proofreading skills of my mother, Sharon Bousum, and the support of my father, Kenneth Bousum, I would have really struggled through this project. I owe a special thanks to my husband, Donald Marcotte, for putting up with me during one of the most stressful times of my life. I am eternally grateful for his enduring love and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

How do scores on the Draw A Person: A Quantitative Scoring System (DAP-Q) compare to scores on the Stanford Binet Intelligence Test: Fourth Edition (SB:FE) in predicting school achievement? If the DAP-Q is comparable to the SB:FE in predicting school achievement, it can be used as a simple, time-saving assessment tool for educators of pre-school and kindergarten children.

Early detection of learning problems and prediction of school readiness is essential in getting children the education they deserve. Drawing tests are not threatening to children because they only have to draw a picture of a person - a task they most likely have performed many times. If these tests can successfully predict school achievement, they will prove useful in detecting children who are behind, so that these children can get the special attention and services they need.

The use of children's drawings to assess their cognitive functioning has a long history. Several different tests have been used to assess children's drawings, including the Bender-Gestalt, the Draw a Person Test, and the Piaget House-Tree Drawing Task (Tramontana, Hooper & Selzer, 1988). Most

measures of intellectual functioning currently in use, including the Stanford-Binet Intelligence Scales, the McCarthy Scales of Children's Abilities and the Wechsler Intelligence Scales for Children, include drawing tasks.

Goodenough (1926) believed there were cognitive elements in children's drawings: "Drawing is not just a visual image but the child draws what he knows" (p. 72). Goodenough (1926) developed the Draw a Man test. Harris (1963) revised Goodenough's test and further developed it as a measure of intelligence. Harris called this revision the Goodenough-Harris Draw a Person. Harris' revision (1963) included a draw a woman and a draw a self along with the draw a man. The purpose of the Goodenough-Harris Draw a Person Test is to provide a simple, time-saving device for measuring intellectual maturity (Reisman & Yamokoski, 1973).

Recently a further refinement of scoring the Goodenough-Harris has been nationally standardized and published, Draw A Person: A Quantitative Scoring System (DAP-Q; Naglieri, 1988). The current study will examine the DAP-Q and the SB:FE to see if the two tests are similarly useful for predicting school achievement.

The DAP-Q is a relatively new instrument, therefore research on it is limited. Most of the literature reviewed focuses on prediction of school achievement based on Harris' 1963 revision (DAP). Kraemer and Tomes (in-review) found that the DAP and the DAP-Q showed no significant differences in predicting SB:FE scores. However, incorporating the

DAP-Q's self-drawing scores improved the DAP-Q's overall predictive utility. The contribution of the current study is to establish how well this new revision, the DAP-Q, predicts later school achievement.

One of the difficulties with previous scoring systems for the Draw a Person was the significant difference found between the scores of boys and girls. Girls were found to score significantly higher than boys on both the Goodenough and the Harris revision. Harris (1963) found the sex differences so significant that separate scoring norms were devised. Naglieri (1988) found the sex differences on the DAP-Q non-significant, therefore separate scoring norms were not devised. Although no differences emerged between boys and girls in her study, Gottling (1990) suggests that any study involving the Draw a Person needs to examine gender differences. The current study of the DAP-Q will examine gender differences in the children's scores in predicting school achievement.

Some researchers (Flynn & Flynn, 1978) believe that there is a problem in predicting later achievement using kindergarten-aged children. These researchers suggest that the difficulty is due to the child's developmental variability and not the specific test. Children's development in the five to six year age range fluctuates; therefore their rates of development may not have a strong relationship with later achievement in school. This researcher intends to look at the predictive validity for both age groups (five-year-old

group and seven-year-old group) on the Stanford-Binet and the DAP-Q.

This comparison of predictive validity between the DAP-Q and the SB:FE yields several possible research questions. Can the DAP-Q predict school achievement as well as the SB:FE? Does the DAP-Q predict school achievement better for girls than for boys? Do seven-year-old's scores on the DAP-Q predict school achievement better than five-year-old's scores? These questions are important in many ways. If the DAP-Q can predict school achievement as well as the SB:FE, it may be a good screening tool for educators of young children. If the DAP-Q predicts better for girls than for boys, it may only be useful for girls. If the DAP-Q predicts better for older rather than younger children, it may not be beneficial to use it with young children.

CHAPTER II

LITERATURE REVIEW

Research by Goodenough (1926), Harris (1963), and Naglieri (1988), supports the use of drawing tests as good measures of intellectual maturity in children. The original test and its revisions were referred to in the literature by a myriad of names. For the purposes of this paper, clarity is achieved through references to the original Goodenough (1926) as the Draw a Man (DAM), the Harris (1963) revision as the Draw a Person (DAP), and the Naglieri (1988) revision as the Draw a Person: A Quantitative Scoring System (DAP-Q).

Why Use a Drawing Test?

Shipp and Loudon (1964) contend that preschoolers' activities are limited to simple things like drawing and oral instructions. Therefore a drawing test would seem to be a simple, efficient measure of a child's intellectual development. The main purpose of human figure drawing tests is to provide a relatively simple, time-saving assessment tool that can be used to estimate intelligence (Reisman & Yamokoski, 1973). Human figure drawing tests are good screening devices because of their ease in administration, scoring efficiency and wide acceptance by children of all ages (Goldman &

Velasco, 1980). The Draw a Person is easily administered and functions as a rapport builder for examiners (Tramill, Edwards & Tramill, 1980).

Knowledge of concepts has been found to be directly related to human figure drawings. Because concepts can be taught, a child's exposure to environmental and educational experiences could and probably does influence his success on human figure drawings (Gottling, 1990).

Hilgert and Adams (1989) state that research on the psychodiagnostic use of children's drawings has been weak. They suggest that research be aimed at finding out which aspects of the DAP-Q are valuable and how they can be improved and employed for greater usefulness. Some researchers question the validity of drawing tests as measures of conceptual maturity or intelligence (Scott, 1981; Barrett, 1983). There is much debate among researchers as to how valid these drawing tests and their scoring systems actually are.

Drawing Development Occurring in Stages

Piaget adopted the principle of children drawing what they know and outlined stages in children's drawings that correspond with his theory of children's development (Piaget & Inhelder, 1969). According to Piagetian theory there are identifiable sequential stages of a child's intellectual development. The emerging major systems are language, mental imagery, symbolic play and drawing. Piaget views drawing as

characterized by "imitative accomodation" in which the child adjusts his drawings to make them represent reality more accurately. Harris (1963) identified three stages in children's drawings. The first, very early, stage consists of the child's pleasure in just making marks on paper. The second stage includes imitative and reproductive drawings. The third, and more developmentally advanced stage, includes the child considering balance, design, and arrangement in his drawing to make it better represent the real world.

Chappell and Steitz (1993) investigated the age-stage relationship between children's human figure drawings and Piaget's levels of cognitive development. These researchers found that as cognitive ability increased so did drawing level. Chappell and Steitz (1993) suggest that children's human figure drawings can be used as a simple measure of cognitive levels in young children. Bensur and Eliot's (1993) research found evidence that developmental changes in children's drawing can reliably demonstrate changes in intellectual development.

Intelligence Tests as Predictors of School Achievement

Tests such as the Wechsler Intelligence Scale for Children (WISC-R), the Kaufman Assessment Battery for Children (K-ABC), and the Stanford-Binet Intelligence Scale: Fourth Edition (SB:FE) are commonly used to assess intelligence in children. Although definitions of

intelligence may vary, the aforementioned tests are accepted as measuring the same constructs.

Research by Laurent, Swerdlik and Ryburn (1992) report that the SB:FE provides a measure of general mental ability as good as any other widely used intelligence test. Concurrent validity of the SB:FE is demonstrated by a correlation coefficient of .89 between SB:FE Composite score and both the Mental Processing Composite and the Achievement score of the K-ABC (Nuttall, Romero, & Kalesnik, 1992). A correlation coefficient of .77 was found between the SB:FE Composite score and the WISC-R Composite score, also demonstrating SB:FE concurrent validity.

Laurent, Swerdlik, and Ryburn (1992) believe that tests of intelligence should also be able to predict school achievement. Much research (Nuttall, Romero & Kalesnik, 1992; Laurent, Swerdlik, & Ryburn, 1992; Tramontana, Hooper & Selzer, 1988) shows the SB:FE to correlate highly with school achievement tests such as the Woodcock-Johnson and the Metropolitan Readiness Test. As well as a measure of intelligence and school achievement prediction, many clinicians have used the Stanford-Binet as a clinical interview (Anastasi, 1988).

Research on Drawing Tests Predicting School Achievement

Shipp and Loudon (1964) found correlations of .51 between the DAM IQ total scores and total achievement scores.

These researchers concluded that the DAM had some value as a predictor of achievement in the first grade and was as good a predictor as other measures of intelligence.

In one study by Duffey, Ritter and Fedner (1976), the Developmental Test of Visual-Motor Integration and the DAM were found to be significant predictors of academic success in the second grade. However, each measure accounted for only 9.3% of the variance and thus were found to have little predictive utility.

The DAP's predictive utility was measured using kindergarten children. Flynn and Flynn (1978) tested the predictive validity of the Slosson Intelligence Test, Peabody Picture Vocabulary Test, the Draw a Person, the Developmental Test of Visual Motor Integration, and the Metropolitan Readiness Test. These researchers found that only the Metropolitan Readiness Test was a significant predictor of school achievement, and even this test accounted for a mere 10% of the total variance of the test. Koppitz (1968) reported that the Bender-Gestalt, a test of mental development, along with tests of human figure drawings was a better predictor of first grade achievement than either measure alone. It is reported, in one study, that the DAP accounted for less than 10% of the variance in measures of academic achievement (Scott, 1981). A study by Serwer, Shapiro and Shapiro (1972) found that kindergarten teachers' ratings were the best predictors of first grade achievement.

Scott (1981) found that the DAP was a reliable measure for children between the ages of 5 and 12, but disputes its validity as a predictor of academic achievement. He reports the DAP accounting for only 10% of the variance on measures of achievement.

The Gesell Institute in Connecticut has developed several tests used to determine school readiness (Ilg & Ames, 1972). Ilg and Ames state that the Incomplete Man Test portion of their Gesell School Readiness Test is the most highly predictive of all the measures. These researchers also state that this part of the test is inevitably the child's favorite. These researchers write of a teacher who, when faced with 37 new pupils on the first day of school, gave them the Incomplete Man Test and later found that she had placed all but two students correctly based solely on the results of this one portion of the test.

In studies reviewed by Tramontana, Hooper, and Selzer (1988), IQ tests such as the Wechsler Preschool and Primary Scale of Intelligence, the Stanford-Binet, and the Slosson were the best predictors of later achievement. Predictive utility of perceptual motor tests including the Bender-Gestalt, Developmental Test of Visual-Motor Integration, and the Draw a Person, contributed effectively to the prediction of reading, math, and general achievement at least through the first grade. However, Tramontana, et al. (1988), concluded that there was not a single measure or set of

measures that invariably provided an accurate prediction of a child's academic success.

In a comparison of the SB:FE, DAP, and DAP-Q, Kraemer and Tomes (in-review) found that both the DAP and the DAP-Q tapped only a small part of the whole which makes up intelligence as measured by the SB:FE. These researchers found that the DAP-Q self score was the only significant correlate and predictor of the SB:FE Composite score. Kraemer and Tomes (in-review) do not recommend that the DAP-Q be used in place of the SB:FE.

Research on Gender Differences in Drawing Ability

Studies by Egan (1986) suggest that girls are significantly ahead of boys in drawing skills. Girls are more likely to copy a square earlier than boys. At 48 months of age, 53% of girls can copy a square compared to only 31% of boys.

Both Goodenough (1926) and Harris (1963) found sex differences in their drawing tests on the drawing of a man. These differences were even greater in the drawing of a woman; girls did better than boys by 3 to 6 raw score points (Harris, 1963; Scott, 1981). Goodenough (1926) termed the differences she found between boys and girls as qualitative and did not find it necessary to devise separate norms.

Shipp and Loudon (1964) report higher scores for girls on the DAM as well as on scores of total achievement. However, the DAM predicted equally well for boys and girls.

Harris' (1963) revision included a draw a woman along with the draw a man. He found significant differences between the performance of boys and girls; therefore, separate norms were created.

Mortensen (1984) argued that the sex differences in the scores on drawing tests cannot be ignored. He believed that these sex differences clearly show that drawing is not a purely cognitive activity. Scott (1981) believes that reported differences may reflect real differences in ability, in which case separate norms only ignore differences that should be further investigated.

Naglieri (1988) and Gottling (1990) reported no significant differences between boys and girls, although girls consistently scored higher than boys. Naglieri (1988) found the sex differences non-significant on the DAP-Q and separate scoring norms were not devised.

Although no differences emerged between boys and girls in her study, Gottling (1990) suggests that any study involving the DAP-Q needs to examine gender differences.

Prediction Problems for Younger Children

Flynn and Flynn (1978) state that predicting later achievement for kindergarten children is a difficult process because of the differing rates of development demonstrated by

children at that age. Other researchers (Shipp & Loudon, 1964) have also observed difficulties in attempts to measure and predict school achievement.

Serwer, Shapiro and Shapiro (1972) found that the DAP was useless in predicting first grade achievement. These authors concluded that the correlation between the DAP & the Metropolitan Achievement Test was insignificant.

Summary and Hypotheses

The current study will look at how scores on the Draw A Person: A Quantitative Scoring System (DAP-Q) compare to scores on the Stanford Binet Intelligence Test: Fourth Edition (SB:FE) in predicting school achievement. The validity of the SB:FE as a widely used, accepted measure of intelligence and a predictor of school achievement has been clearly established in the literature. This test provides a comparison base against which the predictive utility of the DAP-Q may be measured.

As mentioned, sex differences are a key concern among researchers looking at human figure drawing tests. Even though Naglieri (1988) does not report significant sex differences on the DAP-Q, the current study will look at sex differences in the DAP-Q's ability to predict school achievement.

Goldman and Velasco (1980) and Flynn and Flynn (1978) report differences in the predictive utility of human figure drawing tests for children of different ages. These

researchers seemed to support the view that scores obtained from children ages five and six were not useful for predicting school achievement. The current study will examine differences in predictability for two age groups, ages five and seven.

The current study will address the following hypotheses relating to children's scores on the DAP-Q and SB:FE and school achievement prediction.

H_1 Taken as a group, the children's standard scores on the Draw A Person: A Quantitative Scoring System will significantly predict their scores on the Iowa Test of Basic Skills, the school's standardized achievement test.

H_2 Taken as a group, the children's standard scores on the Stanford-Binet Fourth Edition will significantly predict their scores on the Iowa Test of Basic Skills, the school's standardized achievement test.

In addition to testing the above hypotheses, we will also assess the relative contribution of the SB:FE and the DAP-Q in the prediction of the children's school achievement.

H_3 The children's standard scores obtained on the Draw A Person: A Quantitative Scoring System for the seven-year-old group will predict school achievement, as measured by the Iowa Test of Basic Skills, better than the scores for the five-year-old group.

H_4 Taken as a group, the children's standard scores on the Draw A Person: A Quantitative Scoring System will predict school achievement, as measured by the Iowa Test of Basic Skills, better for girls than for boys.

CHAPTER III

METHODOLOGY

The data on the DAP-Q and the SB:FE used in the current study were collected originally for another study titled "Children's Picture Drawing, Cognitive Functioning and Neuromotor Development" (Tomes & Heilbuth, 1991). The students' scores on the school's achievement test are the original data that were collected for this study.

Subjects

The original sample consisted of 72 normal children, divided into four subject groups: 18 five-year-old boys (mean age = 5 yr 5 mo; range is 5-3 to 5-11), 18 five-year-old girls (mean age = 5 yr 4 mo; range is 5-2 to 5-11), 21 seven-year-old boys (mean age = 7 yr 3 mo; range is 7-0 to 7-11), and 15 seven-year-old girls (mean age = 7 yr 4 mo; range is 7-1 to 7-11). All 72 children in the original study were asked to participate in the current study. Of the 72 children in the original study, only 34 agreed to participate in the current study. The sample collected for the current study consisted of 34 normal children, divided into four subject groups: 4 five-year-old boys (mean age = 5 yr 7 mo; range is 5-4 to 5-9), 6 five-year-old girls (mean age = 5 yr

6 mo; range is 5-3 to 5-11), 16 seven-year-old boys (mean age = 7 yr 4 mo; range is 7-1 to 7-11), and 8 seven-year-old girls (mean age 7 yr 5 mo; range is 7-1 to 7-10). The sample population consisted of predominantly white children of varying socioeconomic backgrounds ranging from lower- to upper-middle-class. The children were students from a Public School System, in a small midwestern town in Oklahoma. The seven-year-old subjects were recruited from two public elementary schools. Most of the five-year-old subjects attended half-day kindergarten programs and the remaining were from a half-day preschool program. Participation was on a voluntary basis.

Measurements

The Draw A Person: A Quantitative Scoring System (DAP-Q) is a revision of the DAP. This revision by Jack Naglieri (1988) includes 4 scores: A Man, Woman, Self and a Total score. This is an effort to modify and overcome statistical weaknesses in the DAP scoring system. New norms were created to reflect the updated United States population. Each drawing has a maximum of 64 points. A standard score is derived for each drawing and for the total with a mean of 100 and a standard deviation of 15 (Naglieri, 1988). Reliability coefficients are high (.86), higher than those of the previous two tests. Validity was similar to the DAM and the DAP.

The DAP-Q has a time limit of five minutes per drawing. The total time for this test is 15 minutes: five minutes for each drawing (man, woman, self).

Mental functioning was measured using the Stanford Binet Intelligence Scale: Fourth Edition (SB:FE; Thorndike, Hagen & Sattler, 1986). The latest revision (Thorndike, Hagen & Sattler, 1986) is designed for children 2 years to young adulthood. It is composed of 15 subtests (Nuttall, Romero & Kalesnik, 1992).

Standardized in 1985, the sample reflects the 1980 US Census statistics. Reliability is reported in the high (.95 to .97) range. The modern Stanford-Binet has distinctive patterns of mean scores for Hispanics, Asians, and Blacks - indicating that this test may be racially biased (Kaplan & Saccuzzo, 1989).

The SB:FE yields a composite IQ score and four area scores: Verbal Reasoning, Abstract/Visual Reasoning, Quantitive Reasoning, and Short-Term Memory. Fifteen subtests are distributed among the four areas. Composite IQ and factor scores have means of 100 and standard deviations of 16. The SB:FE takes approximately one hour to 90 minutes to administer.

School achievement was measured using the Iowa Tests of Basic Skills (ITBS). The scores for this test were collected during 1992 and 1993. The subjects' mean ages at the time of achievement testing were: Five-year-old boys, mean age was 6 years 7 months, five-year-old girls, mean age was 6 years 7

months, seven-year-old boys, mean age was 7 years 7 months, and seven-year-old girls, mean age was 7 years 11 months. The ITBS is a battery of tests that measures vocabulary, reading, writing, study skills, listening and mathematics abilities. Reliabilities for the ITBS range from .71 to .91 at the kindergarten and first-grade levels. The Word Analysis, Mathematics and Reading subtests are the only subtests at the kindergarten and first grade level whose reliabilities are high enough (.80) to be used in screening children (Salvia & Ysseldyke, 1991). The ITBS Total Composite National Percentiles are the achievement scores that were used in the current study. Development and standardization of the test seems exemplary (Salvia & Ysseldyke, 1991).

Procedures

In the original study (Tomes & Heilbuth, 1991) letters of intent and a description of the purpose and methods of the research project were mailed to the parents of the subjects.

A battery of six tests was administered to each subject:

1) The Piaget House-Tree Drawing Task
2) The Goodenough-Harris Draw a Person Task
3) The Stanford-Binet Intelligence Scale
4) The Nebraska-Wisconsin Cognitive Assessment Battery
5) The Developmental Test of Visual-Motor Integration
6) The McCarron Assessment of Neuromuscular Development

Testing spanned a one month period. Testing was divided into three sessions each lasting approximately one hour. Testing for the original study took place during the fall of 1991 (Tomes & Heilbuth, 1991). The results of the

Goodenough-Harris Draw a Person test were rescored using the Draw a Person: A Quantitative Scoring System (Naglieri, 1988). The DAP-Q has a five minute time-limit for each drawing. The test administrators report that no student took longer than five minutes to complete each drawing, however no time limits were imposed. Examiners attempted to accomodate each child's schedule, and no children were tested through lunch or recess (Tomes & Heilbuth, 1991).

The children were tested during the 1992-93 school year on the school's yearly standardized achievement test, ITBS. These results were collected on written permission of the proper school authorities and the parents of the children. On request, a report of the results of the research project will be disseminated to the parents. All recording and reporting of data is by subject number. Strict confidentiality has been and will continue to be maintained.

Statistical Methods

A series of hierarchical regression analyses were conducted. In the first regression, the ITBS scores were the outcome and SB:FE and DAP-Q scores were the predictors. In order to assess the relative contributions of DAP-Q and SB:FE, the sizes of the $\underline{R}^2$'s for the DAP-Q and SB:FE were compared. In the second regression, ITBS scores were the outcome and age, gender and DAP-Q scores as well as the interaction between age and DAP-Q and gender and DAP-Q were the predictors. In order to carry out this regression

analysis, gender (male=0; female=1) and age (five-year-olds=0; seven-year-olds=1) were converted to dummy variables.

This researcher expected the results of this study to be similar to the findings of past research. It was expected that the DAP-Q, with its improved scoring system, would be a better predictor of school achievement than its predecessors. This researcher also expected that girls scores would predict school achievement significantly better than boys. Significance was computed at the .05 and the .01 levels. As supported in the literature, this researcher believed that scores for the seven-year-old group would predict school achievement significantly better than scores for the five-year-old group.

CHAPTER IV

RESULTS OF THE STUDY

A table of raw scores by subject number is presented in Appendix A. Means and standard deviations are presented in a table in Appendix B. Prior to conducting regression analyses, Pearson Product Moment Correlations were calculated among all variables. The correlation matrix is reproduced as Table I.

TABLE I

CORRELATION COEFFICIENTS FOR SB:FE, DAPQ, ITBS, AGE AND GENDER

	AGE	GENDER	SB:FE	DAPQ	ITBS
AGE	1.00				
GENDER	-.25	1.00			
SB:FE	.09	-.26	1.00		
DAPQ	-.03	.13	.35*	1.00	
ITBS	-.02	-.17	.63***	.35*	1.00

* $p<.05$; ** $p<.01$, *** $p<.001$

The correlation coefficient for DAP-Q and SB:FE was significant (.35, $p<.05$). Also significant was the correlation between the DAP-Q and the ITBS (.35, $p<.05$). The SB:FE correlated significantly with the ITBS scores (.63, $p<.001$)

Hypothesis 1

To test hypothesis one, a hierarchical regression analysis was conducted. The criterion variable was ITBS percentile. On the first block of the regression, age and gender were entered. On the second block of the regression, DAP-Q scores were entered. Results of this regression analysis are presented in Table II.

TABLE II

HIERARCHICAL REGRESSION ANALYSIS OF AGE, GENDER AND DAP-Q ON ITBS SCORES

Variable	r^2	r^2 Change	F-value	Sig. of F
AGE,GENDER	.03	.03	.54	.59
DAPQ	.17	.14	5.12	.03*

* $p<.05$; ** $p<.01$, *** $p<.001$

As can be seen from Table II, hypothesis one is supported. After the effects for gender and age are removed, the DAP-Q explained 14% of the variance in ITBS scores. This amount of variance is considered significant ($p<.05$).

Hypothesis 2

To test hypothesis two, a hierarchical regression analysis was conducted. The criterion variable was ITBS percentile. On the first block of the regression, age and gender were entered. On the second block of the regression, SB:FE scores were entered. Results are presented in Table III.

TABLE III

HIERARCHICAL REGRESSION ANALYSIS OF AGE, GENDER AND SB:FE ON ITBS SCORES

Variable	r^2	r^2 Change	F-value	Sig. of F
AGE,GENDER	.03	.03	.54	.59
SB:FE	.41	.37	18.84	.0001***

* $p<.05$; ** $p<.01$, *** $p<.001$

As can be seen from Table III, hypothesis two is supported. After the effects for gender and age were removed, the SB:FE explained 37% of the variance in ITBS

scores. This amount of variance is considered significant ($p<.001$)

Hypothesis 3

To test hypothesis three, a hierarchical regression analysis was conducted. The criterion variable was the ITBS percentile. On the third block of the regression, the interaction between age and DAP-Q were entered. Results of this regression analysis are presented in Table IV. As can be seen from Table IV, hypothesis three is not supported by this set of data. The interaction effects of age and DAP-Q did not explain a significant amount of variance in ITBS scores. Therefore, seven-year-olds scores on the DAP-Q did not predict school achievement as measured by the ITBS significantly better than the DAP-Q scores for the five-year-old group.

TABLE IV

INTERACTION EFFECTS OF AGE AND DAPQ ON ITBS SCORES

Variable	r^2	r^2 Change	F-value	Sig. of F
AGE/DAPQ	.18	.00	.14	.70
GENDER/DAPQ	.19	.02	.54	.47

* $p<.05$; ** $p<.01$, *** $p<.001$

Hypothesis 4

To test hypothesis four, a hierarchical regression analysis was conducted. The criterion variable was ITBS percentile. On the third block of the regression, the interaction between gender and DAP-Q were entered. Results of this regression analysis are presented in Table IV on page 25.

As can be seen from Table IV, hypothesis four is not supported by this set of data. The interaction effects of gender and DAP-Q did not predict a significant amount of the variance in ITBS scores. Therefore, girls scores on the DAP-Q did not predict school achievement as measured by the ITBS significantly better than did the boys DAP-Q scores.

Summary

The correlation matrix found three significant correlations. The DAP-Q correlated with the SB:FE significantly (r=.35, p<.05), the DAP-Q correlated significantly with the ITBS (r=.35, p<.05), and the SB:FE correlated significantly with the ITBS (r=.63, p<.001).

Of the four hypotheses predicted, only two found some support. The DAP-Q was found to predict 14% of the variance in ITBS scores (F=5.12, p<.05). Also the SB:FE was found to predict 37% of the variance in ITBS scores (F=18.84, p<.001).

CHAPTER V

DISCUSSION

Summary

The purpose of this study was to see how well the DAP-Q compared to the SB:FE in predicting school achievement. These measures are relatively new and research on their predictive validity is limited. Hypotheses for this study were based on literature which indicated that the Draw a Person and the Stanford-Binet should be able to predict school achievement adequately. The results of this study showed that both measures predicted school achievement significantly, thus hypothesis 1 and 2 were supported. Using hierarchical regression analyses, the results of this study showed that the SB:FE explained 37% of the variance in achievement scores, and that the DAP-Q explained 14% of the variance in achievement scores. According to Cohen's (1977) standards for effect size, regression coefficients accounting for 13 to 26% of the variance in the dependent variable represent medium effects. Regression coefficients accounting for 26% or more of the variance in a dependent variable represent large effects. According to Cohen's (1977) standards, the DAP-Q explained a medium amount of variance in ITBS scores, and the SB:FE explained a large amount of variance. The

literature also indicated that older children's scores on intelligence tests tended to predict their school achievement better than scores for younger children. The results of this study did not find that this was so, thus hypothesis 3 was not supported. Also indicated in the literature was that girls scores on drawing tests were generally better than were scores for boys. This study hypothesized that girls scores on the DAP-Q would predict school achievement better than would boys scores on the DAP-Q. The results of this study found no support for hypothesis 4.

The SB:FE predicted ITBS scores significantly (F=18.84, p<.001), and the DAP-Q predicted ITBS scores significantly (F=5.12, p<.05). This suggests that the DAP-Q, SB:FE and the ITBS are measuring similar constructs.

Conclusions

It seems that although the DAP-Q predicts school achievement significantly, the SB:FE predicts school achievement better. A reason for this may be found in the verbal loading of tasks on the SB:FE and on the ITBS. Both the SB:FE and the ITBS have tasks that require adequate verbal skills. The ITBS has no subtest which measures abstract-visual thinking or short term memory capability. These are suggested by Kraemer and Tomes (in-review) as being the major constructs which the DAP-Q taps in intelligence tests such as the SB:FE. Perhaps if achievement tests such

as the ITBS had such subtests, the DAP-Q would be an even better predictor of school achievement.

It seems obvious from this study that as was suggested by Goodenough (1926), Harris (1963), Naglieri (1988) and Gottling (1990) drawing tests such as the DAP-Q do measure cognitive ability in children. This research seems to show, however, that it may not measure cognitive ability as well as do verbally loaded measures of intelligence such as the SB:FE.

This study may have limited generalizability and results will be interpreted with caution. The current study should be repeated, implementing the five-minute time limit required by the DAP-Q. Although examiners reported that no student required more than five minutes, the DAP-Q is standardized with a five-minute time limit. The sample collected in the original study by Tomes and Heilbuth (1991) are from a small geographic area, this is a second limitation of the current study. This study used data from a relatively small sample size (n=34). The SB:FE's and the DAP-Q's predictive utility may be increased when a larger sample is used. The small sample size also resulted in even smaller samples within groups. This may help explain why gender and age had no significant predictive utility in ITBS scores. If the size of the groups were increased, statistical power would be increased.

This study contributes to the literature on predictive validity of both the SB:FE and the DAP-Q. Both are

relatively new instruments and research on them should be continued. It is recommended that future studies be done to replicate the findings in this study, thus strengthening this study's reliability and validity. Future research on the DAP-Q might look at its usefulness with special populations such as children with dyslexia or other language disabilities. Possible future research for the DAP-Q could also include finding out how the DAP-Q's predictive utility can be improved.

REFERENCES

Anastasi, A. (1988). Psychological testing (6th ed.). New York: Macmillan.

Ayer, F. C. (1916). The psychology of drawing. Baltimore, MD: Warwick & York.

Barret, M. D. (1983). The study of children's drawings: Piagetian and experimental approaches. Early Child Development and Care, 12, 19-28.

Bensur, B. & Eliot, J. (1993). Case's developmental model and children's drawings. Perceptual and Motor Skills, 76, 371-375.

Chappell, P. A. & Steitz, J. A. (1993). Young children's human figure drawings and cognitive development. Perceptual and Motor Skills, 76, 611-617.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences (rev. ed.). New York: Academic Press.

Duffey, J. B., Ritter, D. R. & Fedner, M. (1976). Developmental test of visual-motor integration and the Goodenough draw-a-man test as predictors of academic success. Perceptual and Motor Skills, 43, 543-546.

Egan, D. F. & Brown, R. (1986). Developmental assessment: 18 months to 4 1/2 years. Performance tests. Child: care, health and development, 12, 339-349.

Eng, H. K. (1931). The psychology of children's drawings from the first stroke to the coloured drawing. London. Kegan Paul, Trench, Trubner Co. Ltd.

Flynn, T. & Flynn, L. A. (1978). Evaluation of the predictive ability of five screening measures administered during kindergarten. Journal of Experimental Education, 46, 65-69.

Goldman, R. & Velasco, M. (1980). Toward the development of a rational scale in the use of human-figure drawings as a kindergarten screening measure. Perceptual and Motor Skills, 50, 571-577.

Goodenough, F. L. (1926). Measurement of intelligence by drawings (rev. ed.). Chicago, IL: World Book.

Gottling, S. H. (1990) The validity of the "Draw A Person: A Quantitative Scoring System" for kindergarten children. Unpublished Doctoral Dissertation, Ohio State University, Columbus, Ohio.

Harris, D. B. (1963). Children's drawings as measures of intellectual maturity. New York: Harcourt, Brace & World.

Hilgert, L. D. & Adams, W. F. (1989). Using the Bender-Gestalt test to predict graphomotor dimensions of the Draw-A-Person test. Perceptual and Motor Skills, 68, 27-32.

Ilg, F. L. & Ames, L. B. (1972). School readiness; behavior tests used at the Gesell Institute. New York: Harper & Row.

Kaplan, R. M. & Saccuzzo, D. P. (1989). Psychological testing: Principles, applications, and issues. California: Brooks/Cole.

Koppitz, E. M. (1968). Psychological evaluation of children's human figure drawings. New York: Grune & Stratton.

Kraemer, S. & Tomes, R. (In-Review). Comparison of the Stanford-Binet Intelligence Scale with the Goodenough-Harris Draw a Person and the Draw a Person: A Quantitative Scoring System.

Laurent, J., Swerdlik, M., and Ryburn, M. (1992). Review of Validity Research on the Stanford-Binet Intelligence Scale: Fourth Edition. Psychological Assessment, 4(1), 102-112.

Mortensen, K. (1984). Children's human figure drawings, Vol. 1. Denmark: Dansk Psykologisk Forlag.

Naglieri, J. A. (1988). Draw a person: A quantitative scoring system. New York: The Psychological Corporation.

Nuttall, E. V., Romero, I. & Kalesnik, J. (1992). Assessing and screening preschoolers: Psychological and educational dimensions. Boston: Allyn & Bacon.

Piaget, J. & Inhelder, B. (1969). The psychology of the child. New York: Basic Books.

Reisman, J. M. & Yamokoski, T. (1973). Can intelligence be estimated from drawings of a man? Journal of School Psychology, 11(3), 239-243.

Salvia, J. & Ysseldyke, J. (1991). Assessment, 5th Ed. Boston, MA: Houghton Mifflin.

Sattler, J. M. (1974). Assessment of children's intelligence. Philadelphia: Saunders.

Scott, L. H. (1981). Measuring intelligence with the Goodenough-Harris drawing test. Psychological Bulletin, 89, 483-505.

Serwer, B., Shapiro, B, & Shapiro, P. (1972). Achievement prediction of "high-risk" children. Perceptual and Motor Skills, 35, 347-354.

Shipp, D. E. & Loudon, M. L. (1964). The Draw-a-Man test and achievement in the first grade. The Journal of Educational Research, 57(10), 518-519.

Terman, & Merrill. (1973). Stanford-Binet intelligence scale: 1972 norms edition. Boston: Houghton/Mifflin.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). The Stanford-Binet Intelligence Scale, Fourth Edition-technical manual. Chicago, IL: Riverside.

Tomes, R. & Heilbuth, L. (1991). Children's picture drawing, cognitive functioning, and neuromotor development. Unpublished data. Research funded by grants from the OSU Office of the Assistant Vice President for Research, the College of Education, and the College of Home Economics.

Tramill, J. L., Edwards, R. P., & Tramill, J. K. (1980). Comparison of the Goodenough-Harris drawing test and the WISC-R for children experiencing academic difficulties. Perceptual and Motor Skills, 50, 543-546.

Tramontana, M. G., Hooper, S. R. & Selzer, C. S. (1988). Research on the preschool prediction of later academic achievement: A review. Developmental Review, 8, 89-146.

APPENDIX A

RAW DATA BY SUBJECT NUMBER

#[a]	AGE[b]	GENDER[c]	SB:FE	DAP-Q	ITBS
102	2	2	113	117	52
103	2	2	088	085	45
104	2	2	093	061	28
106	2	2	101	102	41
110	2	2	103	086	98
112	2	2	069	091	24
114	2	2	094	074	86
115	2	2	091	071	37
201	2	1	086	072	30
203	2	1	102	082	74
204	2	1	106	066	34
205	2	1	106	114	91
206	2	1	100	057	34
207	2	1	112	101	89
209	2	1	084	072	46
211	2	1	094	099	76
212	2	1	102	083	63
213	2	1	090	088	43
214	2	1	104	090	65
216	2	1	098	089	58
217	2	1	110	095	91
218	2	1	093	081	16
219	2	1	103	071	74
221	2	1	096	106	43
306	1	2	104	081	66
307	1	2	085	078	32
312	1	2	095	104	26
314	1	2	079	073	07
315	1	2	096	109	70
318	1	2	101	103	99
412	1	1	102	073	97
414	1	1	087	075	51
415	1	1	090	074	40
417	1	1	115	096	80

[a] Subject Number
[b] Age (1 = 5 yr olds, 2 = 7 yr olds)
[c] Gender (1 = boys, 2 = girls)

APPENDIX B

MEANS AND STANDARD DEVIATIONS FOR THE SB:FE, DAP-Q, AND ITBS BY GROUP

	n	SB:FE		DAP-Q		ITBS	
		M	SD	M	SD	M	SD
BOYS	20	99.00	8.84	84.20	14.63	59.75	23.69
GIRLS	14	93.71	11.17	88.21	16.57	50.79	28.87
5 YR	10	95.40	10.60	86.60	14.65	56.80	30.83
7 YR	24	97.42	9.99	85.54	15.92	55.75	24.31
TOTAL	34	96.82	10.05	85.85	15.34	56.05	25.91

VITA

Debra F. Marcotte

Candidate for the Degree of

Master of Science

Thesis: A COMPARISON BETWEEN THE DRAW A PERSON AND THE STANFORD-BINET INTELLIGENCE TEST IN PREDICTING SCHOOL ACHIEVEMENT.

Major Field: Family Relations and Child Development

Biographical:

Personal Data: Born in Tulsa, Oklahoma, December 10, 1968, the daughter of Kenneth and Sharon Bousum, and Charles Perkins.

Education: Graduated from Inola High School, Inola, Oklahoma, in May 1987; received Bachelor of Arts Degree in Psychology from University of Tulsa in December, 1992; completed requirements for the Master of Science degree at Oklahoma State University in July, 1994.

Professional Experience: Teaching Assistant, Department of Family Relations and Child Development, Oklahoma State University, August, 1993, to May, 1994.