Assessment Practices of School Psychologists

by

Sarah Rose Whitney

A dissertation submitted to the Graduate Faculty in Educational Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2011

This manuscript has been read and accepted for the
Graduate Faculty in Educational Psychology in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Georgiana Shick Tryon, Ph.D.

Date — Chair of Examining Committee

Mario Kelly, Ed.D.

Date — Executive Officer

Dr. Mary Kopala

Dr. Jay Verkuilen

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

THE CITY UNIVERSITY OF NEW YORK

Abstract

Assessment Practices of School Psychologists

by

Sarah Rose Whitney

Advisor: Georgiana Shick Tryon, Ph.D.

The present study examined the factors related to school psychologists' use of projective tests and what trade-offs practitioners were willing to make between psychometric properties, convenience, and clinical judgment in their choices of hypothetical social-emotional tests. Further, the study explored the variables of professional identity, cognitive dissonance, extraversion, and self-efficacy as they related to practitioners' reported use of projective tests and to their preferences between hypothetical choices. Participants included 116 presenters at the 2010 Annual Convention of the National Association of School Psychologists. Data were collected with an online questionnaire that included measures of cognitive dissonance (Elliott & Devine, 1994; Matz, Hofstedt, & Wood, 2008), extraversion (John & Srivistava, 1999; Matz et al., 2008), and self-efficacy (Chen, Gully, & Eden, 2001; Huber, 2006). Correlational analyses indicated that the explanatory variables sometimes related to practitioners' use of projective tests. Conjoint analysis, using logistic regressions, indicated that when faced with hypothetical test choices, respondents generally showed a preference for the test

with the best psychometric properties, most convenience, and least required clinical judgment. In general, most respondents indicated a preference for hypothetical tests that have robust psychometric properties, but they also reported using actual projective tests in practice. The constructs of professional identity, cognitive dissonance, self-efficacy, and extraversion, and the relationships among those variables and between those and the usage of projective tests provided some explanation for this phenomenon.

Acknowledgments

I am so appreciative of the support and guidance that many people have given me as I worked to complete this study. My advisor, Dr. Georgiana Shick Tryon, has been my mentor in guiding me through the dissertation process. I thank her for her enthusiasm, advice, and encouragement. I thank Dr. Jay Verkuilen for his invaluable help in the data analyses of this study. I am also grateful to Dr. Mary Kopala, Dr. Carol Tittle, and Dr. David Budescu, who graciously gave of their time to assist me.

I owe more thanks than I can say to my family and friends who have cheered me on through the years as I worked through graduate school, and especially through writing this dissertation. My parents and biggest fans, Alma and Richard Whitney, were the first to encourage me to pursue a doctoral degree, and they have been unwavering in their love, support, and humor along the way. My husband, Christopher Harriss, always helped me keep things in perspective, never failed to bolster my confidence, and always kept me laughing. This accomplishment is that much sweeter because I can share it with the people who helped me achieve it. Thank you, thank you, thank you.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1: Introduction

School psychologists (SPs) are experts in assessment (NASP, 2003). Numerous surveys of SPs have found similar results: SPs perform a variety of job tasks, but they spend most of their time conducting assessments (Fagan & Wise, 2007). The results of the assessments that SPs perform often are used in determining educational trajectories for students (Salvia et al., 2007; Tharinger, Pryzwansky, & Miller, 2008).

Because the decisions based on assessments can have long-lasting, profound effects on a child, there exist numerous guidelines and laws for the practice of assessment. The American Psychological Association (APA, 2002), the National Association of School Psychologists (NASP, 2003), the American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME) all have stated guidelines for assessment (AERA, APA, & NCME, 1999); further, chief among these guidelines is the requirement that assessment tools meet high psychometric standards. Laws such as the Individuals with Disabilities Education Improvement Act (IDEIA, 2004) also have explicit regulations for how evaluations of students must be conducted, including using psychometrically sound instruments.

Despite these guidelines and laws, SPs continue to use assessment techniques that arguably do not measure up to the standards set forth in such guidelines and laws. Projective tests in particular would appear not to meet these strict psychometric standards because they often lack norming samples, require considerable clinical judgment in administering and interpreting, and their development is not based on psychometrics, but rather on psychodynamic theory (Kamphaus & Frick, 2005). In addition, researchers have concluded that some of the most commonly used projective measures do not have

adequate psychometric properties (Halperin & McKay, 1998; Lilienfeld, Wood, & Garb, 2000; Sattler & Hoge, 2006). Generally speaking, projective measures do not have strong psychometric properties (Kamphaus & Frick, 2005), yet SPs continue to use them (Hojnoski, Morrison, Brown, & Matthews, 2006; Shapiro & Heick, 2004).

There is no question that the use of projective tests is a controversial issue. There exists much disagreement in the field over whether the use of projective measures is acceptable. Many believe that projective measures do not possess acceptable psychometric properties (Kamphaus & Frick, 2005) and their use is not consistent with best practices (Knoff, 2002). Others, however, welcome the use of projective measures and stress the value of clinical judgment in assessment (Koppitz, 1983). Many practitioners value the use of clinical judgment (Koppitz, 1983) even though it is often inaccurate or incorrect (Bell & Mellor, 2009; Garb, 1989, 2005; Grove et al., 2000; Grove, 2005). It is clear that the use of projective measures is highly controversial, but at the same time, they are not as psychometrically robust as many objective, norm-referenced tests.

This study examined *why* SPs continue to use measures that do not meet criteria set forth by professional and legal guidelines. Perhaps a simple explanation is that projective tests are convenient to administer and score, which contributes to their continued use by practitioners. Indeed, some of the most commonly cited barriers to adherence to guidelines in school psychology and other fields as well are time and resources (Cabana et al., 1999; Kratochwill & Shernoff, 2003).

In addition, the constructs of professional identity, cognitive dissonance, practitioners' personality characteristics and self-efficacy appear to be promising for

studying why SPs continue to use projective measures. SPs begin to develop their professional identities during graduate school (Bruss & Kopala, 1993; Colbeck, 2008), so perhaps students who spend more time on campus interacting with and observing faculty have a better understanding of professional issues, including the necessity to use psychometrically sound assessment tools, than students who spend less time on campus. Cognitive dissonance theory (Festinger, 1957) is another construct that could help to explain why SPs use measures that they know do not meet the standards they are supposed to, or do not meet the standards that other assessment measures do. The field of cognitive dissonance examines, among other things, the processes a person experiences when he or she behaves in a way that is inconsistent with the elements of his or her cognitions (Festinger & Carlsmith, 1959). Certainly cognitive dissonance could easily occur if SPs know that a measure has not been validated psychometrically, know they are only supposed to use psychometrically validated measures, and yet use said measure anyway.

It is also possible that SPs' personalities are related to the types of tests they use in assessment. In particular, research has suggested that the personality characteristic of extraversion is related to feelings of cognitive dissonance; specifically, that extraversion moderates and ameliorates the discomfort that cognitive dissonance causes (Matz, Hofstedt, & Wood, 2008). Self-efficacy, both general and specific, is another construct that could help explain SPs' assessment preferences and test choices. Bandura (1977) defined self-efficacy as "the conviction that one can successfully execute the behavior required to produce the outcomes" (Bandura, 1977, p. 193). That is, perhaps SPs are

convinced that they are able to use projective tests to make desirable recommendations for students.

This study examined the assessment practices and preferences of SPs and the reasons they use various assessment measures. Participants for this research study were volunteer SPs who I solicited by email. An online questionnaire, which I have developed with Dr. Tryon and used in my pilot study, was the primary data-collection measure. The questionnaire contained eight parts, including questions about demographic information, professional identity, social-emotional test preferences, social-emotional test usage, cognitive dissonance, extraversion, general self-efficacy, and SP-specific self-efficacy.

Using the questionnaire, I answered the following research questions:

- Do SPs continue to use projective tests despite lack of psychometric support for projective tests?
- Are SPs' professional identities related to their use of projective tests?
- Are SPs' use of projective tests related to their feelings of cognitive dissonance about doing so?
- Are SPs' use of projective tests related to their ratings of extraversion?
- Are SPs' use of projective measures related to their ratings of general and specific self-efficacy?
- Are SPs willing to accept less psychometric robustness of tests in order to be able to exercise some clinical judgment?
- Do SPs' theoretical orientations relate to their use of projective tests (i.e., projective tests have roots in a psychodynamic theoretical orientation)?

Chapter 2: Literature Review

In this chapter, I will present a review of the literature pertaining to the major topics and constructs that are the focus of this dissertation. I will begin with an overview of the role and function of school psychologists (SPs) and a summary of the professional standards and ethical guidelines from the professional organizations to which SPs belong; in this latter section I will include a discussion of the laws that govern assessment and of evidence-based practices. Then, I will describe objective and projective testing techniques and how they fit into the guidelines. Last, I will discuss barriers to adherence to guidelines, in which I will include a discussion of four constructs (i.e., professional identity, cognitive dissonance, extraversion, and self-efficacy) that may relate to SPs adherence to evidence-based assessment practices. I will end the chapter with the rationale for this study and the hypotheses I will test.

Role and Function of SPs

Definition of a SP

What is an SP? A concise definition is hard to find. Two of the major governing bodies of the practice of school psychology, the National Association of School Psychologists (NASP) and the American Psychological Association (APA), publish lengthy descriptions on their websites concerning the practice of school psychology. On its website, NASP states:

> School psychologists help children and youth succeed academically, socially, and emotionally. They collaborate with educators, parents, and other professionals to create safe, healthy, and supportive learning environments for all students that strengthen connections between home and school. School psychologists are

> highly trained in both psychology and education. They must complete a minimum of a Specialist-level degree program (60 graduate semester credits) that includes a 1200-hour internship and emphasizes preparation in the following: data-based decision making, consultation and collaboration, effective instruction, child development, student diversity and development, school organization, prevention, intervention, mental health, learning styles, behavior, research, and program evaluation. School psychologists must be certified and/or licensed by the state in which they work. They also may be nationally certified by the National School Psychology Certification Board (NSPCB). (NASP, n.d.)

On its website, Division 16 of APA, the Division of School Psychology, states:

> School Psychology is a general practice and health service provider specialty of professional psychology that is concerned with the science and practice of psychology with children, youth, families; learners of all ages; and the schooling process. The basic education and training of school psychologists prepares them to provide a range of psychological assessment, intervention, prevention, health promotion, and program development and evaluation services with a special focus on the developmental processes of children and youth within the context of schools, families, and other systems. School psychologists are prepared to intervene at the individual and system level, and develop, implement, and evaluate preventive programs. In these efforts, they conduct ecologically valid assessments and intervene to promote positive learning environments within which children and youth from diverse backgrounds have equal access to

> effective educational and psychological services to promote healthy development. (APA, n.d.)

Fagan and Wise (2007), leading authorities of the history of SP, offered the following definition:

> A school psychologist is a professional psychological practitioner whose general purpose is to bring a psychological perspective to bear on the problems of educators and the clients educators serve. This perspective is derived from a broad base of training in educational and psychological foundations as well as specialty preparation, resulting in the provision of comprehensive psychological services of a direct and indirect nature. (p. 4)

These definitions are broad, and they reflect the wide range of skills and responsibilities that SPs employ on a daily basis. In essence, SPs use their knowledge of psychology and education to work toward positive outcomes for children, families, and schools.

Major Function of SPs

Historically, assessment has been a major function of the SP (Reschly, 2000). SPs perform a combination of duties including testing, counseling, consultation, and intervention, but they spend the majority of their time conducting assessments (Fagan & Wise, 2007). As long ago as 1938, one scholar noted that, "individual testing has been the most usual function of school psychologists" (Barker, 1938). Numerous surveys since then have corroborated this sentiment. A 1981 survey found that on average, school psychologists spent 47.9% of their time conducting assessments (Goh, Teslow, & Fuller, 1981). Another survey around the same time found that SPs spend 54% of their time on assessments (Smith, 1984). A later replication of the 1981 survey found that SPs spent

half their time conducting assessments (Hutton, Dubes, & Muir, 1992), as did other surveys a few years later (Curtis, Hunley, Walker, & Baker, 1999; Stinnett, Havey, & Oehler-Stinnett, 1994). More recently, Hojnoski et al. (2006) surveyed practicing SPs and found that the modal number of educational assessments they performed during the 2000-2001 school years was 60, and ranged from 4 to 200. Other recent surveys have found that SPs spend between 43% and 46% of their time engaged in assessment (Bramlett et al., 2002; Koonce, 2007). Further, the National Association of School Psychologists (NASP) views SPs as experts in assessments (NASP, 2003). Hence, assessment is a primary function of most SPs.

Definition of assessment. Because SPs spend so much of their time conducting assessments, it is useful to examine what assessment is. Assessment is "a complex problem-solving or information-gathering process" (Fagan & Wise, 2007, p. 117) that is:

> conducted by a psychologist trained to gather a variety of different types of information (review of school history and health records, observations, interviews, and test results) from a number of different sources (pupil, teacher, parents, and specialists) and to interpret or give meaning to that information in light of the unique characteristics of the pupil and his or her situation. (Jacob & Hartshorne, 2007, p. 88)

Whereas testing refers to the "administration and scoring of tests" (NASP, 2003, p. 1), assessment is a multi-modal, multi-method process of gathering data about a student. It is a "process of collecting data for the purpose of making decisions about students" (Salvia, Ysseldyke, & Bolt, 2007, p. 4). Assessment involves a referral process, classroom observations, examination of school records, testing, interviews, report writing, and

conferences to discuss results, recommendations, and further actions (Fagan & Wise, 2007; Sattler, 2008). Distinguishing testing from assessment emphasizes the point that assessment is a comprehensive process that includes testing as one component.

Importance of assessment. In addition to understanding what assessment is, it is useful to understand the purposes for which SPs use assessments. SPs aid in making important educational decisions for students based on the outcomes of the assessments they conduct (Salvia et al., 2007). The "application of the knowledge and skills of school psychologists in assessment can help ensure that all children and youth receive needed services" (NASP, 2003, p. 1). SPs use assessments to investigate difficulties with reading, writing, mathematics, conduct, peer relationships, and internalizing problems such as depression and anxiety (Bramlett et al., 2002). By collecting data through assessment techniques and creating a learning profile of students, SPs are instrumental in implementing interventions such as determining which students have special education needs or necessitate modified instruction (Tharinger, Pryzwansky, & Miller, 2008). Salvia, Ysseldyke, and Bolt (2007) described four categories of decisions based on assessment: prereferral classroom decisions, entitlement decisions, post-entitlement decisions, and accountability/outcome decisions. Each category contains multiple possibilities for outcomes, including providing interventions, classifying a student as needing special education services, or deciding on an appropriate placement for a student (Salvia et al., 2007). Thus, SP's use of assessment techniques can have substantial influence over the educational trajectory of a student by helping to determine the student's most appropriate course of action and educational plan.

Perhaps a striking way to examine the importance of the assessment role of SPs is to review the legal suits that have arisen from the results of such assessments. These legal suits are of two kinds: those in which the assessment procedures are challenged and those in which assessment outcomes are used as evidence (Fagan & Wise, 2007). Many of the historically well-known lawsuits fall into the former category and involve the overrepresentation of cultural minorities in restrictive educational environments, such as special education (see, for example, *Larry P.* v. *Riles,* 1984). To this day, numerous lawsuits arise each year that fall into the latter category; these lawsuits arise over the educational placements that materialize in part due to assessments that SPs perform (see, for example, *T.Y.* v. *New York City Department of Education,* 2009). SP's role in assessment, therefore, cannot be understated; it is underscored by the important educational decisions based on those assessments and by the legal actions that arise in reaction to many of those decisions.

Guidelines for Assessment

The practice of assessment clearly can have a huge impact on those for whom educational decisions are made based on the outcomes of the assessments. As such, it is useful to review the guidelines and laws that govern the practice of assessment.

Professional Guidelines

The professional organizations to which SPs may belong publish guidelines for the specific practice of school psychology (i.e., NASP) and for the practice of psychology in general (i.e., APA). These guidelines include standards for the practice of assessment, specifically for the tests that make up the assessments. For example, NASP endorses using tests that are "scientifically based…technically appropriate and used for the

purposes for which they were developed and/or validated" (NASP, 2003, pp. 1-2). In addition, standard C.1(a) of the NASP professional conduct manual (2000) states:

> School psychologists maintain the highest standard for educational and psychological assessment and direct and indirect interventions. . . . School psychologists are knowledgeable about the validity and reliability of their instruments and techniques, choosing those that have up-to-date standardization data and are applicable and appropriate for the benefit of the child. . . . School psychologists use assessment techniques, counseling and therapy procedures, consultation techniques, and other direct and indirect service methods that the profession considers to be responsible, research-based practice. (pp. 27-28)

Further, standard 9.02(b) of the APA Ethics Code states, "psychologists use assessment instruments whose validity and reliability have been established for use with members of the population tested" (APA, 2002, p. 1071). In other words, SPs are supposed to "use psychological instruments of assessment that have established validity and reliability with populations for which they have been normed" (Nagy, 2005, p. 257). Thus, both APA and NASP have formal guidelines that stress the use of psychometrically robust, research-supported assessment instruments.

In addition, since 1966, APA, in conjunction with the American Educational Research Association (AERA) and the National Council on Measurement in Education (NCME), has published *The Standards for Educational and Psychological Testing* (1999) that details criteria for assessments, including test construction, and is often cited in legal proceedings. For example, Standard 3.1 states, "Tests and testing programs should be developed on a sound scientific basis. Test developers and publishers should compile and

document adequate evidence bearing on test development" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p. 43). Further, these guidelines require that reliability and validity evidence be reported in test materials, and norming populations and procedures must be clearly described.

Legal Guidelines

The laws that govern the practice of assessment in education also stress the use of psychometrically sound instruments. The major law in this respect is the Individuals with Disabilities Education Improvement Act (IDEIA, 2004). This law provides legal regulations for the evaluation of students to determine if they have a disability. Federal regulations to implement Part B of IDEIA provide legal requirements for assessment, stating that whoever conducts the evaluation must "use technically sound instruments" that "are used for the purposes for which the assessments or measures are valid and reliable" (34 C.F.R. §§ 300.304(b)(3), 300.304(c)(1)(iii)). Section 504 of the Rehabilitation Act of 1973, another law that governs much of the practice of SPs, has similar standards for assessment as those in IDEIA. Federal regulations to implement Section 504 also provide legal requirements for assessment, stating that "tests and other evaluation materials have been validated for the specific purpose for which they are used" (34 C.F.R § 104.35(b)(1)). Hence, in addition to the ethical and professional standards described above, legal requirements also exist that necessitate the use of instruments that are reliable and valid.

Evidence-Based Practice

In addition to the APA, NASP, and AERA guidelines and the laws that govern the practice of assessment in schools, best practices (Knoff, 2002) and the trend toward evidence-based practice in the field of psychology stress using empirically supported instruments. Hoagwood and Johnson (2003) defined evidence-based practice (EBP) as:

> a body of scientific knowledge, defined usually by reference to research methods or designs, about a range of service practices (e.g. referral, assessment, case management, therapies, or support services). . . .The knowledge base is usually generated through application of particular inclusions criteria (e.g., type of design, types of outcome assessments) and it generally describes the impact of particular service practices on child, adolescent, or family outcomes. "Evidence-based practice" or EBP is a shorthand term denoting the quality, robustness, or validity of scientific evidence as it is brought to bear on these issues. (p. 5)

Further, Kratochwill and Shernoff (2003) stated that, "the challenge to improve our services to children and school continues, and at the nexus of this challenge is the adoption of research-based (or evidence-based) practices in diagnosis, assessment, and intervention" (p. 390). Hence, it is clear that it is necessary for SPs to employ assessment techniques that adhere to the ethical and practice guidelines, and in doing so, they must employ scientific, evidence-based methods. Indeed, a recent issue of the *Journal of Clinical Child and Adolescent Psychology* emphasized the movement toward EBPs in the assessment of children and adolescents in many areas, including anxiety (Silverman & Ollendick, 2005), depression (Klein, Dougherty, & Olino, 2005), bipolar disorder (Youngstrom, Findling, Kogos Youngstrom, & Calabrese, 2005), Attention Deficit

Hyperactivity Disorder (Pelham, Fabiano, & Massetti, 2005), conduct problems (McMahon & Frick, 2005), and learning disabilities (Ozonoff, Goodlin-Jones, & Solomon, 2005).

In line with utilizing EBPs when conducting assessments, SPs must ensure that the evaluations they conduct are multifaceted, comprehensive, fair, valid, and useful (Jacob & Hartshorne, 2007). Included in the concept of an evaluation being valid is the idea that it is the SP's responsibility to choose measures that "meet high professional standards" and that are appropriate for their assessments, meaning that the measures are reliable, valid and have adequate standardization norms (Jacob & Hartshorne, 2007, p. 96).

How can SPs know whether to use a particular test in their assessment of students or not? Gall, Gall, and Borg (2007) proposed four criteria for judging the quality of tests for use in research, and their statements are equally applicable for SPs when it is their responsibility to choose psychometrically sound instruments. First, a test must be objective, meaning that "its scores are undistorted by biases of the individuals who administer and score it" (Gall et al., 2007, p. 194). As conditions of administration and scoring increase in flexibility, a test decreases in objectivity. This leads to the second condition, which is that tests have standard conditions of administration and scoring "to ensure consistency in administration and scoring across all testing situations" (Gall et al., 2007, p. 195). Third, tests should have standards for interpretation, in the form of criteria or norms. Fourth, a test should be fair, meaning "two groups of equal ability with respect to the construct measured by the test…should earn the same score on each item of the test" (Gall et al., 2007, p. 195).

Psychometric Properties

Because there is so much emphasis on using tests that are reliable, valid, and norm-referenced, I will now discuss each of those concepts. A full discussion of these constructs is outside the scope of this dissertation, but it is important nonetheless to understand what each is because the concepts are used in judging the quality of tests.

Reliability. Reliability refers to consistency of test scores (Salkind, 2006) and how much error is present in a test score (Gall et al., 2007). According to classical test theory, each person has a true score and an error score when taking a particular test, which combine to produce each person's observed score (Salkind, 2006). As the amount of error in scores goes down, the reliability of a test goes up (Gall et al., 2007). Reliability of scores can be measured by alternate-form or parallel forms reliability, test-retest reliability, internal consistency, and interrater reliability (Gall et al., 2007; Salkind 2006). Parallel forms reliability measures how the scores from one form of a test correlate with another form of the same test; test-retest reliability measures how the scores of a test given at one time correlate with the scores of that same test given at a later time; internal consistency reliability measures how the items within the test correlate with the total score; and interrater reliability measures how much agreement there is between raters over the scoring of the test (Salkind, 2006). In essence, reliability refers to the stability and consistency of scores over time, which can be measured in the ways discussed above.

Validity. Validity refers to how closely a test measures what it purports to measure (Salkind, 2006) and whether one can draw meaningful conclusions from the test data (Gall et al., 2007). Validity of a test can be measured by content validity, criterion

validity, and construct validity (Salkind, 2006). Content validity measures how closely the content of a test reflects the "entire universe of items in a certain topic" (Salkind, 2006, p. 66). Criterion validity compares the scores with another test that has already been shown to be valid and assess the same abilities, and looks for a high, positive correlation between the scores on these two tests. Further, there are two types of criterion validity: (a) concurrent validity refers to the correlation with a contemporaneous test, and (b) predictive validity refers to the correlation with a test that will take place sometime in the future (Salkind, 2006). Construct validity measures how closely a test taps into the psychological construct it is supposed to be testing (Salkind, 2006). Therefore, the validity of a measure encompasses how accurately it measures what it is claiming to measure.

How reliable and valid do tests have to be? Salkind (2006) proposes a cutoff of .70 for reliability coefficients in research, but suggests that .80 and above is more acceptable. The standard for clinical, psychoeducational, and screening tests appears to be reliability coefficients higher than .80 and at least .90 for decision-making tests (Salvia et al., 2007; Sattler, 2008). Tests with reliability coefficients between .70 and .79 have moderate reliability, between .60 and .69, have poor reliability, and below .60, very poor reliability (Sattler, 2008). No such specific numerical guidelines are available for deciding how acceptable the validity of a test is (Jacob & Hartshorne, 2007). Instead, to decide if a test is valid, SPs should look at manual and supportive materials, and also at current research about the tests (Jacob & Hartshorne, 2007). It is the test user's responsibility to ensure the valid use of assessment measures (Salvia et al., 2007). In sum, SPs should review supporting materials of and research regarding tests to ensure that

reliability coefficients are at least .90 when they are using the tests to make educational decisions, and that there is robust support for the validity of the test.

Norm-referenced. It is also important to understand what it means for a test to be norm-referenced. When a test uses norm-referenced measurement, it compares each individual's performance on the test with the performance of a specified group (Salvia et al., 2007; Sattler, 2008). Tests that are norm-referenced "allow you to compare one individual's test performance to the test performance of other individuals" (Salkind, 2006, p. 206). Important factors in evaluating the quality of a norm group include its representativeness, its size, and its relevance (Salvia et al., 2007; Sattler, 2008). SPs also should be concerned with the age, grade, gender, geographic region, SES of the norm group, and whether the norms are current (Salvia et al., 2007; Sattler, 2008). When evaluating the size of a norm group, "the sample should contain at least 100 individuals in each age group" (Sattler, 2008, p. 103). When evaluating the relevance of the norm group, SPs need to select the proper norm group, whether it be national or local norms; further, if an SP uses a norm group that most practitioners customarily do not use, then the SP should be clear about which norm group he or she used (Sattler, 2008).

Objective and Projective Testing Techniques

Having established what it means for a test to be psychometrically robust and meet the standards for educational-decision making, I will now turn to a discussion of the classes of objective and projective tests and evaluate how well each group of tests measures up to the standards.

Assessment comprises various forms of testing: cognitive ability tests; academic achievement measures; perceptual tests (visual, auditory); tests of fine and gross motor

abilities; behavioral, personality, and adaptive behavior measures; and curriculum-based assessment (Fagan & Wise, 2007). Thus, the scope of assessment encompasses a multitude of forms of testing various areas of skill and achievement.

Assessment instruments typically fall into one of two categories: objective or projective. Objective tests utilize direct questioning to gather information from participants, whereas projective tests use open-ended stimuli to compel participants to share information they might not otherwise share (Kamphaus & Frick, 2005). Kamphaus and Frick (2005) distinguished objective and projective measures from each other in three ways. First, the results of objective tests are garnered from empirical methods (e.g., comparing the results to a norming sample), whereas the results of projective tests are based on the clinician's interpretation. Second, objective measures are less likely to be based on psychodynamic theory than are projective measures. Third, the development of objective tests is more reliant on psychometrics than is the development of projective tests.

Objective Tests

The category of objective measures includes tests that have standardized administration and scoring procedures as well as national norms and numerous studies supporting their reliability and validity (Kamphaus & Frick, 2005). I will provide an overview of two objective tests below: the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; Wechsler, 2003a), and the Behavior Assessment System for Children – Second Edition (BASC-2; Reynolds & Kamphaus, 2004). The former is an example of a cognitive assessment tool in which responses are marked as correct or incorrect and then scored, and the latter is an example of a personality assessment tool in

which responses mark the presence or absence of various qualities (Sattler & Hoge, 2006).

WISC-IV. The WISC-IV (Wechsler, 2003a) could be considered an objective measure. Its manual contains explicit instructions for administration and scoring individual responses, and scaled scores and index scores are based on the results of a nation-wide norming sample (Wechsler, 2003a). The WISC-IV also is a norm-referenced measure; its norming sample comprises 2,200 children and adolescents stratified according to the 2000 United States Bureau of Census (Wechsler, 2003b). The norming sample was selected to be demographically representative in terms of age, gender, race, geographic region, and parent education (Wechsler, 2003b).

Further, the WISC-IV contains a separate manual that enumerates various psychometric properties of the test and its individual components (Wechsler, 2003b). The WISC-IV technical manual (Wechsler, 2003b) reports reliability and validity data for the test, with reliability coefficients ranging from .85 to .97 for the index scores. The test-retest and internal consistency reliability coefficients, respectively, for the composite scores are: .89 and .94 for Verbal Comprehension, .85 and .92 for Perceptual Reasoning, .85 and .92 for Working Memory, .79 and .88 for Processing Speed, and .89 and .97 for Full Scale IQ (Wechsler, 2003b). The reliabilities of the individual subtests range from .79 to .90 (Wechsler, 2003b). Further, the technical manual (Wechsler, 2003b) contains evidence from numerous validation studies that provide evidence of good criterion and construct validity. Taken together, the psychometric evidence for the WISC-IV suggests that the index scores meet the high standards of measures that SPs use for education-

related decisions (Salvia et al., 2007) and it "is considered to have outstanding reliability" (Sattler, 2008, p. 312).

BASC-2. The BASC-2 (Reynolds & Kamphaus, 2004) is an assessment measure that also could be considered an objective measure. It is a questionnaire format, with forms for Self-Report of Personality (SRP), Teacher Rating Scale (TRS), and Parent Rating Scale (PRS). There are also forms for a Structured Developmental History (SDH) and for Student Observation System (SOS). The SRP, TRS, and PRS measure adaptive behavior, and the respondents' answers produce various scale scores including those for School Problems, Internalizing Problems, Inattention/Hyperactivity, Personal Adjustment, and overall composite scores, the Behavioral Symptoms Index (BSI), and the Emotional Symptoms Index (ESI).

Like the WISC-IV (Wechsler, 2003a), the BASC-2 (Reynolds & Kamphaus, 2004) is a norm-referenced measure; its norming sample comprised 3,400 individuals for the SRP, 4,800 for the PRS, and 4,650 for the TRS, with each age group containing at least 100 participants. Further, the norming sample was selected to be demographically representative of the Current Population Survey of 2001 in terms of age, sex, race, geographic region, parent education, and special education classification (Reynolds & Kamphaus, 2004).

The BASC-2 manual (Reynolds & Kamphaus, 2004) reports reliability and validity data for the test, with internal reliability coefficients above .90 for the BSI, Externalizing Problems, and Adaptive Skills indexes, and above .80 for the School Problems and Internalizing Problems indexes of the TRS. For the PRS, internal reliability coefficients are above .90 for the BSI and Adaptive Skills, and range from about .85 to

about .95 for the Internalizing and Externalizing indexes. For the SRP, internal reliability coefficients are above .90 for the Internalizing Problems and the ESI indexes, and range from the middle to upper .80s for the School Problems, Inattention/Hyperactivity, and Personal Adjustment indexes. The BASC-2 manual (Reynolds & Kamphaus, 2004) also reports test-retest reliabilities of middle .80s to low .90s for the TRS, and in the low .90s for the PRS. One exception for both of these forms is the test-retest reliability of the Internalizing Problems index, for which the test-retest reliability was .78 for both. Test-retest reliabilities for the SRP range from the upper .70s to the low .90s. The BASC-2 manual (Reynolds & Kamphaus, 2004) reports interrater reliability coefficients that are, in general, lower than the coefficients for internal reliability and test-retest reliability. Median values range from .53 to .65 for the TRS, and .69 to .77 for the PRS.

In addition to the extensive reliability information the manual provides, the BASC-2 manual (Reynolds & Kamphaus, 2004) also lists extensive research to support the good validity of the test, including factor analysis to assess construct validity and correlations with other scales to measure criterion validity. The BASC-2 (Reynolds & Kamphaus, 2004) also has built in scales to assess validity: the F index (to measure faking negativity); the L index (to measure faking positivity); the V index (to measure endorsement of nonsensical test items); the Consistency Index (to measure random response patterns); and the Response Pattern Index (to measure if a rater was not paying attention to the content of the items). Taken together, the psychometric evidence for the BASC-2 suggests that many of the indexes meet the high standards of measures that SPs use for education-related decisions, and it is a satisfactory measure to use with school-aged children (Sattler & Hoge, 2006; Salvia et al., 2007).

Projective Tests

In contrast, the category of projective tests includes social-emotional measures for which practitioners typically do not use standardized administration or scoring procedures, and for which typically there is little psychometric evidence (Kamphaus & Frick, 2005). In general, projective tests are of three kinds: a child draws a picture, a child tells a story in response to a picture, or a child interprets an inkblot (Sattler & Hoge, 2006). Certainly, the Thematic Apperception Test (TAT; Murray, 1943) could be considered a projective measure. The practitioner shows the participant a series of black and white pictures and asks the participant to create stories about the pictures. Though some psychologists have devised scoring systems (Kleinmuntz, 1967; McClelland, Atkinson, Clark, & Lowell, 1958) for the TAT, most practitioners do not use the scoring systems that exist (Kamphaus & Frick, 2005; Rossini & Moretti, 1997). As such, the interpretations that result from projective measures rely heavily on the SPs' interpretation of the student's responses that is largely based on clinical judgment and not in reference to a norming sample (Sattler & Hoge, 2006). SPs use of clinical judgment in interpreting projective measures contrasts with the use of objective instruments with sound psychometric properties, especially because clinical judgment often can be inaccurate or incorrect (Bell & Mellor, 2009; Garb, 1989, 2005).

In a review of meta-analyses examining the psychometric properties of projective techniques, Lilienfeld and colleagues (2000) summarized the evidence for three commonly used projective techniques: the Rorschach, TAT, and human figure drawing techniques. First, regarding the Rorschach, they concluded, "despite its continued widespread use by clinicians, the Rorschach Inkblot Test remains a problematic

instrument from a psychometric standpoint" (p. 38). One major problem is that the most commonly used norms for scoring continue to misclassify normal respondents as pathological. In addition, the reliability of many of the variables is unfounded or unknown, and the scientific foundation on which many of the variables are based is weak. Next, in discussing the TAT, the authors concluded that although some scoring systems have been devised, they are not appropriate for routine use because there are inadequate norms and reliability is untested or unknown. Last, the authors concluded that there is little psychometric support for the scores derived from the human figure drawing techniques; however, "perhaps most important, there is no convincing evidence that human figure drawings possess incremental validity above and beyond other readily available demographic or psychometric data" (p. 51). Thus, the result of this review of data suggested that three of the most well known projective techniques lack adequate psychometric properties and/or norms.

Other researchers have found similar results about projective tests. Halperin and McKay (1998) reviewed the Human Figure Drawing Test, House-Tree-Person Test, and Kinetic Family Drawing test and concluded that, "despite their popularity, most projective tests have relatively weak psychometric data supporting their reliability and/or validity" (p. 575). In their review, the authors noted that the drawing techniques should be interpreted with caution. Further, they noted that clinicians rarely use systematic administration on storytelling techniques such as the TAT and Children's Apperception Test (CAT), so "because of the lack of standardized procedures and objectivity in scoring, these results must be interpreted with extreme caution" (p. 583). In sum, Halperin and McKay's (1998) review underscores the notion that objective testing

techniques have better psychometric properties and more valid and reliable results than do projective testing techniques.

Sattler and Hoge (2006) noted that the Draw-A-Person Test (DAP), in which a child simply has to draw a person and tell a story about the person, has no standardized norms, "reliability is weak because of the variety of scoring procedures used," (p. 292), and validity also is weak. Thus, SPs should not use it as an assessment technique, but perhaps as a clinical tool instead (Sattler & Hoge, 2006). The authors also reviewed the Children's Apperception Test (CAT; Bellak & Bellak, 1949), an adaptation of the TAT in which children make up stories about a series of 10 drawings of animals in human-like situations. The manual has no standard scoring procedures, norms, reliability data, or validity data; therefore, it is impossible to judge the psychometric properties of the CAT (Sattler & Hoge, 2006). Sattler and Hoge (2006) also reviewed the Rorschach and concluded that the measure administered with the Exner scoring system "must be used cautiously with children" (p. 294).

In general, projective measures do not have strong psychometric properties (Kamphaus & Frick, 2005), yet SPs continue to use them. Recent survey research shows that school psychologists still utilize projective measures in assessments (Hojnoski, Morrison, Brown, & Matthews, 2006; Shapiro & Heick, 2004). For example, Hojnoski and colleagues (2006) surveyed 500 practicing SPs from the 2002 NASP member list. The 175 respondents estimated the number of assessments, both educational and social-emotional, that they conducted during the previous academic year, the number of the social-emotional assessments based on projectives, and indicated which instruments were of primary importance in the testing. They also ranked each instrument on its usefulness,

and reported how frequently they used a measure for specific educational purposes. In general, the majority of the sample reported using projective measures in assessment; for example, 60% of the sample reported using a sentence completion measure. In sum, SPs are using projective measures to make important educational decisions, a use for which the projective techniques are not validated.

Numerous other studies have illustrated the continued use of projective measures. Watkins, Campbell, Nieberding, and Hallmark (1995) surveyed 412 clinical psychologists to examine their most frequently used assessment measures. Of the top 10 most frequently used tests, four were projectives (Sentence Completion Methods, TAT, Rorschach, Projective Drawings). Other studies have found similar results, that is, that psychologists in various settings utilize projective measures with some frequency (Camara, Nathan, & Puente, 2000; Cashel, 2002; Sweeney, Clarkin, & Fitzgibbon, 1987; Vukovich, 1983). SPs also use projective tests on a regular basis. Prout (1983) surveyed 300 school psychologists (practitioners and trainers) about their social-emotional assessment methods and found that practitioners rated behavioral observation and clinical interview as more important and more utilized than projective tests; projective measures dominated the top 10 most frequently used social-emotional assessment techniques reported. A later study (Wilson & Reschly, 1996) surveyed 251 SPs as to their most frequently used instruments in assessment and found that three of the top ten most used were projective tests. More recently, Miller and Jome (2008) found that many SPs think projective measures are important to use in assessing internalizing disorders.

Further, SPs' use of projectives in certainly is not a new phenomenon. A national survey in 1981 investigated 274 SPs' usage of psychological tests (Goh, Teslow, &

Fuller). Intellectual, academic, and perceptual-motor assessment ranked highest in number of measures given for each area. In addition, measures with well-established psychometric properties were favored in the most frequently used measures, but in personality assessment, respondents listed more projective measures (e.g., sentence completion and House-Tree-Person). Further, in personality assessment, the most frequently used measures were projective ones.

The use of projective techniques is a highly controversial and long-debated issue. Thelen, Varble, and Johnson (1968) surveyed 140 clinical psychology faculty members and found much disagreement over whether projective tests are supported by research and over whether learning projective tests should be required coursework. Pruitt, Smith, Thelen, and Lubin (1985) later replicated this study with 153 respondents and found even more disagreement between faculty members. Some argue that when evaluated in terms of psychometric properties, projective techniques usually do not measure up (Kamphaus & Frick, 2005), but perhaps could be used as a self-exploration technique in therapy rather than as an assessment measure (Garb, Wood, Lilienfeld, & Nezworski, 2002). Many practitioners agree that the use of projective tests in assessing children is not consistent with best practices (Knoff, 2002), especially when there are other more reliable, valid measures that rely less on clinical judgment (Miller & Nickerson, 2006, 2007). Sattler and Hoge (2006) "recommend that important decisions about individuals never be made solely on the basis of their performance on projective measures" (p. 291). Further, the use of projective measures opens up the possibility for litigation against a school for using a projective technique as part of an educational decision (Peterson & Batsche, 1983). Last, the clinical judgment required to administer and interpret projective

measures often can be inaccurate or incorrect (Bell & Mellor, 2009; Garb, 1989, 2005); research has shown that clinical judgment rarely is more accurate than statistical predictions of human behavior, psychological diagnoses or prognoses, or personality (Grove et al., 2000; Grove, 2005).

Some practitioners, however, are adamant about the use of projectives and the importance of clinical judgment in assessment. Koppitz (1983) argued that "the use and validity of drawings in the assessment of school children is well established despite often inconclusive research results. . . . The value derived from children's drawings depends on the knowledge, experience, and skill of the examiner" (p. 426). "Projective techniques have long been, are, and, we suspect, will always be among the assessment procedures most frequently used by clinical psychologists. . . . Their place in clinical assessment now seems as strong as, if not stronger than, ever" (Watkins, Campbell, Nieberding, & Hallmark, 1995, p. 59). Others are less certain about a definitive answer concerning the use of projective tests, stating "there can be no absolute answer about whether or not to use projectives with school children. Concerns about the validity and usefulness of personality tests, like other assessment tools, are appropriately addressed by considering test properties in relation to the purposes of the assessment" (Jacob & Hartshorne, 2007, p. 110). It is clear that the use of projective measures is highly controversial, but at the same time, they are not as psychometrically robust as many objective, norm-referenced tests.

Barriers to Adherence to Guidelines

Having established the criteria for selecting assessment measures and the fact that SPs continue to use measures that do not meet those criteria, it is important to understand

why this is the case. To date, there are no studies of this phenomenon as it applies to SPs. One study examined test use by geographic region in the United States (Hosp & Reschly, 2002) and found that SPs in New York, New Jersey, and Pennsylvania administered more projective tests than SPs in other parts of the country, but the study did not delve into the *why* of the usage. In their writing on EBPs, Kratochwill and Shernoff (2003) noted that one challenge to using evidence-based interventions in practice is that procedures may contradict practitioners' beliefs or theoretical orientations. Further, the authors stated that "even when psychologists are aware of the empirical evidence supporting a technique or procedure, they may not infuse this evidence into practice because doing so would require more work than time permits or more resources than are available" (Kratochwill & Shernoff, 2003, p. 392). Thus, if a test is inconvenient practitioners may choose other tests with less evidence to back them up simply because they are more convenient to use. This convenience factor is not unique to the field of school psychology: Cabana et al. (1999) found that many physicians do not adhere to practice guidelines because of decreased time and increased costs.

Research in other fields of psychology has suggested possible reasons why psychologists do not follow practice guidelines. Studying the field of clinical psychology, Medeiros (2002) noted that adherence to clinical practice guidelines may be hindered by out-of-date guidelines; questions about the relevance and usefulness of a particular guideline to a particular client; poor implementation and dissemination of guidelines; lack of time to locate best practices and critically review them; disagreement between experts about benefits of alternatives; desire to adhere to current therapeutic methods and beliefs of other clinicians who are viewed as experts (but who may not necessarily be

following guidelines); and research evidence that may be incomplete or inconclusive (Medeiros, 2002).

Given the ethical and practice guidelines that govern the practice of school psychology, and the field's emphasis on EBPs, it is important to consider why SPs continue to use projective measures. In the next section, I will discuss four constructs that may help to explain SPs' behavior in this regard. First, I will discuss professional identity. Second, I will describe cognitive dissonance. Third, I will describe the personality trait of extraversion. Fourth, I will discuss self-efficacy. I will conclude by proposing how measuring these variables may contribute to the understanding of why SPs behave the way they do. Though a full discussion of each of these theories is beyond the scope of this dissertation, I will provide a summary of each of the theories and of the relevant research.

Professional Identity

It is possible that a SP's sense of professional identity affects how closely one follows the practice guidelines and ethical standards, specifically with the use of projective measures. VanZandt (1990) defines professionalism, thereby also defining professional identity, as:

> the way in which a person relies on a personal high standard of competence in providing professional services, the means by which a person promotes or maintains the image of the profession, a person's willingness to pursue professional development opportunities that will continue to improve skills within the profession, the pursuit of quality and ideals within the profession, and a person's sense of pride about the profession. (pp. 243-244)

For example, belonging to professional organizations reflects a sense of professional identity in that it reflects in a practitioner the desire to be up-to-date with the issues and practices within the profession (VanZandt, 1990).

Graduate school is a formative time for professional identity. Bruss and Kopala (1993) suggest that "graduate school training in psychology may be viewed in terms of a 'professional infancy' wherein an individual enters the field with limited professional awareness, skills and understanding, and an undeveloped sense of professional identity" (p. 686). Students' progressions through graduate school reflect growth not unlike that an individual experiences from infancy through adulthood, and "it is the task of the faculty and training staff to nurture and promote growth, in a role similar to that of parents" (Bruss & Kopala, 1993, p. 686). Thus, graduate school is a crucial time for professional identity development. It is during this time that "students learn their chosen profession's abstract body of professional knowledge and its associated skills during lengthy degree programs and apprenticeships" and they "observe the behaviors, attitudes, and norms for social interaction prevalent among practitioners of their profession" (Colbeck, 2008, p. 9). SP faculty members presumably are more current in their understanding of the field by virtue of their scholarly activities than are practitioners in the field. Thus, students who spend more time on campus interacting with and observing faculty should have a better understanding of professional issues such as the necessity to use psychometrically sound assessment tools than students who spend less time on campus.

Cognitive Dissonance

Cognitive dissonance theory. Cognitive dissonance theory (Festinger, 1957) might help to explain why SPs use measures that they know do not meet the standards

they are supposed to, or do not meet the standards that other assessment measures do. Cognitive dissonance is "an unpleasant state of tension generated when a person has two or more cognitions… that are inconsistent" (Oskamp, 1984, p.401) or when a person behaves in a way that is inconsistent with the elements of one's cognitions (Festinger & Carlsmith, 1959).

Festinger (1957) proposed that people strive for consistency within themselves, between "what a person knows or believes and what he does" (p. 1). When such consistency is unavailable or impossible, for example when a person continues to smoke despite knowing the health risks of doing so, cognitive inconsistency exists. This inconsistency is what Festinger (1957) termed dissonance, and he described the various ways in which cognitive dissonance can arise. First, "dissonance could arise from logical inconsistency" (Festinger, 1957, p. 14). To illustrate this, Festinger (1957) gave the example of a person who believed that humans could travel to the moon but could not build the device necessary to do so. Second, "dissonance could arise because of cultural mores" (Festinger, 1957, p. 14). To illustrate this possibility, Festinger (1957) provided an example of a person eating with his hands at a formal gathering. Third, "dissonance may arise because one specific opinion is sometimes included, by definition, in a more general opinion" (Festinger, 1957, p. 14). For this point, Festinger (1957) described the possibility of a person favoring a candidate from a rival political party. Fourth, and last, "dissonance may arise because of past experience" (Festinger, 1957, p. 14). Festinger (1957) explained this scenario by describing a person standing in the rain but not getting wet. Festinger (1957) postulated that it is unlikely that any person would never experience any cognitive dissonance, but instead suggested, "it is necessary to distinguish

degrees of dissonance and to specify what determines how strong a given dissonant relation is" (p. 16). He referred to this as the magnitude of dissonance and suggested that it is influenced by how important the cognitive elements are that are creating the dissonance.

Further, Festinger's theory of cognitive dissonance (1957) suggests that people strive to reduce or eliminate cognitive dissonance, which acts as motivation. Festinger postulated that cognitive dissonance is an uncomfortable state of being and that people strive to reduce the feelings associated with the dissonance (Stone, 2001). In addition, the greater the magnitude of the cognitive dissonance, the stronger the drive to reduce it (Festinger, 1957). People can reduce dissonance by changing the way they behave or feel, changing the environment that creates the dissonance, or adding new cognitive elements to reconcile dissonant elements (Festinger, 1957). Despite being motivated and even taking action to reduce dissonance, there is no "guarantee that dissonance will be reduced" (Festinger, 1957, p. 23). In order to reduce cognitive dissonance, people may have to make a change that could be painful or involve a loss; they may be happy with their present behavior; or changing the behavior may just be impossible (Festinger, 1957).

Cognitive dissonance research. Cognitive dissonance is a broad theory that can be applied in many different ways. A database search yields studies that have used cognitive dissonance to study expatriates' experiences with cultural change (Maertz, Hassan, & Magnusson, 2009), eating disorders (Rodriguez, Marchand, Ng, & Stice, 2008), experiences of people climbing Mount Everest (Burke, Sparkes, & Allen-Collinson, 2008), finance (Olsen, 2008), online gaming (Chiou & Wan, 2007), drug use

(Peretti-Watel, 2006), and terrorism (Maikovich, 2005). Thus, cognitive dissonance is a far-reaching way of thinking about cognitions and behavior that is relevant to many disciplines, and such a wide-reaching theory is promising for studying psychological practice. Despite the wealth of studies examining cognitive dissonance, however, none has studied the construct in relation to SPs or assessment practices. Further, there are no studies examining cognitive dissonance as it relates to psychologists, teachers, or educators in general. As such, I will review recent literature of what circumstances cause dissonance and how dissonance influences actions and attitudes. This is the most relevant area of research, because this dissertation will examine assessment practices of SPs as they relate to feelings of cognitive dissonance. Certainly a state of tension, or cognitive dissonance, can occur if SPs know that a measure has not been validated psychometrically, know they are only supposed to use psychometrically validated measures, and yet use said measure anyway.

A recent study (Starzyk et al., 2009) examined whether the cognitive dissonance reduction strategies that people use varies depending on how important and salient their attitudes are to them. The authors conducted their study with 308 undergraduate students who had to write essays conveying an argument either for or against government-supported student loans. Prior to writing, the students expressed their actual thoughts on the topic and rated the importance of their views, which were grouped into low, moderate, and high importance. Participants then were randomly assigned to testing conditions based on choice (if they had a choice to write for or against student loans) and salience (whether they were reminded of their own personal views before writing the

piece). After writing a piece in contrast with their actual beliefs, the students again rated their views on the topic and the importance of their views.

The results of this study show that attitude change varied as a function of importance and salience. The experimenters found that low and moderate importance participants dealt with dissonance by altering their attitudes. High importance participants, however, changed their attitudes and engaged in trivialization in order to reduce dissonance. So, participants in the high-importance, low-salience condition restructured their cognitions the most. The study also found that when people were keenly aware of their views (high salience), and those views were important to them, it was harder for them to restructure their cognitions to reduce dissonance. Perhaps SPs who engage in activities that contradict best practices have views that are important to them for various reasons, and this makes it difficult for them to change their views and/or behavior.

In another study, McKimmie and colleagues (2003) studied whether group salience affects the need to reduce cognitive dissonance. Participants were 99 undergraduate students at a university in the Netherlands. Participants were randomly assigned to six testing conditions based on the combinations of behavioral support (support or nonsupport) and group salience (individual identity, low group salience, or high group salience). The experimenters collected data using a questionnaire, and they told participants that they were studying peoples' views of personality qualities, specifically generosity. Participants in the high salience group had to write about how they were similar to other students at their university but different from students at another university in the same city; these participants also were told that the purpose of

the study was to see how students from their university compared to the other students. Participants in the low salience group had to write about what they had in common with other students from their university, but there was no writing about the contrast with students from another university. Participants in the individual identity group had to write three positive things about themselves.

After the self-description task, all groups participated in an hypocrisy manipulation in which they first rated how important generosity is as an attribute. They also completed a Generosity Checklist (measuring donations to charity, paying for friends' drinks, and giving money to homeless people) that the experimenters used to make participants aware of their own past ungenerous behavior. Participants then completed a measure of psychological discomfort, which gauged how comfortable or nervous they were feeling after completing the previous measures. To manipulate the behavioral support condition, the experimenters gave the participants information from a fake study conducted at their university examining how often the students performed the activities on the Generosity Checklist. In the nonsupport condition, the fake results showed that at least 80% of students engaged in each of the activities on the checklist, and in the support condition, the fake results showed that fewer than 25% of students engaged in each of the activities on the checklist. Lastly, the participants in the real study completed measures of attitude change and group identification.

The results of this study suggested that lack of salient group support can lead to increased attitude change and decreased group identification. Though there was no effect on attitude change as a function of behavioral support in the individual identity group, participants in the high salience group showed a much larger reversal of attitudes when

they were also in the nonsupport group. In addition, for participants in the high salience group, higher levels of group identification were associated with behavioral support. This study is important because it illustrates the social context of cognitive dissonance. Participants who experienced the most dissonance distanced themselves from the salient group, in effect, to reduce their dissonance. This study is also relevant to the study of SPs' behavior because it suggests that practitioners could be affected and influenced by what their peers and co-workers are doing if they feel an affinity toward them and support from them.

Elliot and Devine (1994) used induced compliance to examine the psychological discomfort associated with cognitive dissonance. The authors had as participants several hundred college students from the University of Wisconsin-Madison. The participants completed a questionnaire measuring their attitudes about many issues relevant to their school, for example, the idea that the school should raise tuition. Participants who believed strongly that the school should not increase tuition were then selected to participate further in the study, part of which was writing an essay either for or against the tuition increase. The participants did not have a choice of whether they had to write the essay in support of or against the tuition increase. This formed the forced compliance component of the study. Participants also completed an affect measure, an attitude change measure, and a choice manipulation and attitude check to see how much choice the participants believed they had in writing for or against the issue. The affect measure consisted of 24 items related to feelings of dissonance, for example, "uncomfortable." The participants rated on a 7-point Likert scale to what extent they were feeling certain emotions while participating in the study. The authors averaged the responses to

"uncomfortable," "uneasy," and "bothered" to calculate a discomfort score. The remaining items were used to measure negative and positive emotions. The results of the this study suggested that people who must express opinions that are counter to their beliefs experience greater feelings of cognitive dissonance than those who can express opinions that are in line with their beliefs.

Other studies have adapted and used the measure introduced by Elliot and Devine (1994). Matz and Wood (2005) and Matz, Hofstedt, and Wood (2008) also used the emotion measure to assess cognitive dissonance, positive emotions, and negative self-evaluations. In the latter study, the authors investigated whether personality factors can serve as mediators for feelings of cognitive dissonance. The authors interviewed 188 undergraduate students in Texas to examine attitude change and psychological discomfort related to dissonance. The participants first completed a short attitude survey and then were told or not told (depending on the experimental condition) that they would be meeting with other participants to discuss one of the issues on the survey. Further, the students were led to believe that the other group members either agreed or disagreed with their own positions. The students completed an attitude survey, feedback forms about the other group members, an emotion measure, a measure of final attitude, and an extraversion measure. The emotion measure was replicated from the Elliot and Devine (1994) study. The participants rated on a 7-point Likert scale the extent that they were feeling, among other things, "uneasy," "uncomfortable," "tense," "bothered," and "concerned" at the moment; these five items were used to determine the cognitive dissonance score of each participant. The results of this study will be discussed further below in the discussion of extraversion.

Cognitive dissonance measures. Investigators typically have measured cognitive dissonance using self-report methods. Participants in the McKimmie et al. (2003) study rated on a 9-point Likert scale how comfortable, relaxed, calm and nervous they were feeling. Participants in the Elliott and Devine (1994) study rated on a 7-point Likert scale to what extent they were feeling certain emotions while participating in the study. The authors averaged the responses to "uncomfortable," "uneasy," and "bothered" to calculate a discomfort score. Last, participants in the Matz and Wood (2005) and Matz et al. (2008) studies responded on 7-point Likert scale the extent that they were feeling, among other things, "uneasy," "uncomfortable," "tense," "bothered," and "concerned" at the moment; these five items were used to determine the cognitive dissonance score of each participant.

Personality Characteristics – Extraversion

Perhaps SPs' personalities also relate to their use of projective tests. Studies have examined the relationship of demographic characteristics and SPs' professional practices (Curtis, Hunley, & Grier, 2002; Curtis, Hunley, Walker, & Baker, 1999; Vukovich, 1983), but these studies did not examine the use of projective tests or personality characteristics in particular.

Development of the Big Five. It appears that the first time a set of five distinct personality factors was proposed was in 1949. Drawing on the work of earlier personality theorists, Fiske (1949) collected data from first-year clinical psychology trainees from six groups of 24 students. Each group spent one week at an assessment center, during which time they lived together. Also during this time, the participants completed various tests and biographical inventories, and trainers closely observed the trainees. At the end of the

week, each trainee rated himself, his fellow trainees, and the trainers rated the trainees as well. They used a rating scale with 42 bipolar personality variables (e.g., predictable *v.* unpredictable). The authors found the median reliability of the trainer ratings was .84 and of the trainee ratings was .52. Using factor analysis, the authors identified five recurrent factors: social adaptability, emotional control, conformity, the inquiring intellect, and confident self-expression. Though the names of these factors would change with further research, as I will describe below, this study appears as one of the first mentions of five distinct factors of personality.

Tupes and Christal (1961, reprinted in 1992) also identified five major personality traits. The purpose of their study was to "isolate meaningful and relatively independent trait-rating factors which are universal enough to appear in a variety of samples" (p. 227). The authors used a sample of 1,816 participants from the following groups: Air Force officer candidates, graduates of officer candidate school, students in Air Force command school, undergraduate students, and graduate students. Each participant rated his or her peers in each group on 35 bipolar (e.g., silent *v.* talkative, depressed *v.* cheerful) personality characteristics. Psychologists and psychiatrists also rated each of the participants on the same characteristics. The results of this study underscored the idea of five strong factors of personality. From the ratings in this study, five rotated factors emerged. The factors that Tupes and Christal (1961) identified were: surgency (talkative, frank, adventurous, assertive, sociable, energetic, composed, interested in opposite sex, and cheerful); agreeableness (good-natured, not jealous, emotionally mature, mild, cooperative, trustful, adaptable, kind, attentive to people, and self-sufficient); dependability (orderly, responsible, conscientious, perseverant, conventional); emotional

stability (not neurotic, placid, poised, not hypochondriacal, calm, emotionally stable, and self-sufficient); and culture (cultured, esthetically fastidious, imaginative, socially polished, independent-minded).

Norman (1963) replicated the five factor model of personality traits. Participants were 622 students at the University of Michigan who completed peer nomination ratings of each other. The author's factor analysis of the peer nominations yielded five factors of personality: extroversion (this spelling later changed to extraversion, but it is not clear when or why that change took place) or surgency (talkative, frank, adventurous, sociable), agreeableness (goodnatured, not jealous, mild/gentle, cooperative), conscientiousness (fussy/tidy, responsible, scrupulous, persevering), emotional stability (poised, calm, composed, not hypochondriacal), and culture (artistically sensitive, intellectual, polished/refined, imaginative).

The five factors of personality were largely ignored in research and publication during the 1960s and 1970s, but enjoyed a rebirth in the 1980s (McCrae & John, 1992). Several studies then confirmed again the existence of five robust categories of personality traits, and researchers began referring to them as the Big Five (Digman, 1989; Goldberg, 1981; John, Angleitner, & Ostendorf, 1988). Digman (1989) reviewed multiple meta-analyses and listed the Big Five as they are known today: extraversion, agreeableness, conscientiousness, neuroticism, and openness.

Definition of extraversion. As described above, the specific definitions of each of the Big Five personality traits have changed since their original stating, but the underlying concept of the trait of extraversion has remained consistent (Matz, Hofstedt, & Wood, 2008). Some of the adjectives used to describe the extraversion trait are active,

assertive, energetic, enthusiastic, outgoing, and talkative (McCrae & John, 1992), and some of the adjectives to describe the other end of the spectrum, the introversion trait, are reserved, socially aloof, and less interpersonally effective (Matz et al., 2008).

In addition, there is some evidence that the biological characteristics of extraverts may be different than those of introverts. Eysenck (1990) reviewed studies that used electroencephalography (EEG), electrodermal responses, salivation, and enzyme measures to examine possible differences between introverts and extraverts. Results of the many studies he culled suggested that extraverts may possess a less excitable nervous system than introverts (Eysenck, 1990); therefore, extraverts may be less likely to experience negative feelings associated with arousal (Matz et al., 2008). Matz and colleagues (2008) hypothesized that, "if extraverts are less prone than introverts to experience the psychological discomfort associated with arousal, then extraverts may find dissonance-producing situations less unpleasant than introverts" (p. 402). Within cognitive dissonance investigations, researchers have studied various phenomena to explain behavior and behavior change associated with cognitive dissonance. Hence, the trait of extraversion may be useful in exploring the ways in which people deal with cognitive dissonance.

Personality research. There exists some literature focusing on SPs and personality factors; particularly, studies have examined personality factors and job burnout (Mills & Huebner, 1998), personality factors in general (Davis & Sandoval, (1992) and personality factors of award winning psychologists (Toomey, Levinson, & Morrison, 2008). Mills and Huebner (1998) studied the personality and environmental factors that contribute to burnout for SPs. Specifically, they sought to determine if any of

the five factors of the Big Five as particular personality factors contributed to burnout. Participants were SPs registered with the North Carolina Department of Public Instruction who completed a demographic questionnaire, the School Psychologists and Stress Inventory (SPSI; an occupational stressors scale), the NEO Five-Factor Inventory (NEO-FFI; a personality inventory), and the Maslach Burnout Inventory (MBI; a burnout scale). The SPSI lists possible stressful scenarios and asks participants to rate each on how much of a problem each scenario has been in their jobs. Mills and Huebner (1998) did not report specific psychometric data for the SPSI, but noted that it underwent "an extensive content validation procedure" (p. 108) and "additional validity data were provided" (p. 109) in a previous study. The NEO-FFI assesses the Big Five personality traits: extraversion, neuroticism, openness to experience, agreeableness, and conscientiousness. Each trait has its own domain of 12 questions, and the authors report coefficient alphas from .74 to .89 for the domain scales. The MBI measures occupational well-being with three scales: Emotional Exhaustion (measures how emotionally overextended participants feel), Depersonalization (measures how much of an insensitive or dehumanized attitude participants have), and Reduced Personal Accomplishment (measures participants' feelings of competence and success). Participants rate each of the items in these scales on how often they experience the feelings, ranging from never to every day. The authors reported internal consistency coefficients of ".90 for Emotional Exhaustion, .79 for Depersonalization, and .71 for Reduced Personal Accomplishment" (p. 108).

Though the authors reported many complex findings about personality characteristics and burnout, I will focus on the results related to extraversion because

those are the most relevant to this dissertation. Mills and Huebner (1998) reported results at two times, first to report cross-sectional results, and second to report longitudinal results. The cross-sectional results and the longitudinal results indicated that extraversion was negatively correlated with Emotional Exhaustion and positively correlated with Reduced Personal Accomplishment. Mills and Huebner (1998) suggested that "school psychologists who report high emotional exhaustion and reduced personal accomplishment often display tendencies toward introverted behavioral responses (e.g., withdrawn, passive, reserved reactions)" (p. 115). This study is important because it illustrates how personality characteristics, and extraversion in particular, can be related to and affect how SPs carry out and feel about their jobs.

A recent study (Toomey et al., 2008) examined the vocational personality traits of school psychologists. A sample of 241 NASP members completed a data form that assessed demographic information and The Self-Directed Search Revised (SDS-R), a personality measure used for career counseling. Though a full discussion of the theory behind the SDS-R is beyond the scope of this paper, it is important to note that it is based on Holland's Theory of Vocational and Work Environments, which the authors noted is "arguably the most well accepted and empirically tested theory of vocational personalities" (Toomey et al., 2008, p. 418). The theory proposes six personality types: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional; based on the results of a person's completion of the SDS-R, a person's personality type can be characterized by a three letter code, where each letter is the first of one of the traits (Toomey et al., 2008). The SDS-R has four scales: Activities (to measure likes and dislikes), Competencies (to measure abilities), Occupations (to measure feelings towards

various vocations), and Self-Estimates (to measure estimations of abilities) (Toomey et al., 2008). The authors reported internal consistency coefficients from .90 to .94 and test-retest reliability coefficients from .76 to .89 for the scales.

The results of this study indicated a personality type for SPs that is dominant in the Social trait. The authors noted that this is an expected result because SPs are engaged in a helping profession, and people with "social personalities enjoy and thrive in situations that afford them the opportunity to utilize their skills in interpersonal communication, social problem solving, and counseling" (Toomey et al., 2008, p. 422). This study found that SPs also have strong Enterprising and Investigative traits, which might be reflected in their positions to persuade teachers and administrators and use science as a guide for selecting practices (Toomey et al., 2008). As the authors noted, there was a possibility for response bias in their study, but it is the first study that characterized the vocational personality of SPs.

Davis and Sandoval (1992) also investigated personality characteristics of SPs and contrasted personalities of award-winning SPs with those of a random sample. The authors collected data from 100 members of the California Association of School Psychologists; they also included 74 SPs who had received that organization's Outstanding Psychologist award for the previous 16 years. Participants completed a demographic questionnaire and the California Personality Inventory (CPI), which researchers have used "in a variety of other fields to predict training and vocational outcomes" (Davis & Sandoval, 1992, p. 418), but which had not to this point been studied with SPs. Davis and Sandoval (1992) did not report specific psychometric properties of the CPI, but instead noted that it had recently been renormed and has "good

psychometric properties" (p. 418). Results of the CPI yield scales including Empathy, Psychological Mindedness, and Achievement via Conformance.

Davis and Sandoval (1992) found that both groups of SPs scored high on measures of Self-Realization, Dominance, Independence, Achievement via Conformance, Achievement via Independence, and Psychological Mindedness, which suggests that SPs are "confident, resourceful, desire to do well and achieve, are interested in why people do what they do…and tend to be good judges of how people feel" (p.419). Award winners did differ from the random sample on measures of Capacity for Status, Good Impression, and Achievement via Conformance, which suggests that the award winners could be more ambitious, more inclined to perform better when tasks are well defined, and yet eager to please others (Davis & Sandoval, 1992). The authors cautioned against overinterpreting these results because the differences between the groups never exceed four scaled score points, although the differences were statistically significant. Nonetheless, this is an interesting study that examines the personality traits of SPs, and of a specific sample of award winners.

Another study that is perhaps most relevant to this dissertation examined extraversion as it relates to cognitive dissonance. In the attitude change study described in the section above, Matz et al. (2008) examined the personality trait of extraversion as a moderator of discomfort caused by cognitive dissonance. The experiment found that extraversion moderates the psychological discomfort that cognitive dissonance can cause. In general, participants who believed that other group members agreed with their views felt less dissonance discomfort than those who believed the others disagreed with them. When examined in more detail, however, specifically by looking at personality traits, the

extraverts did not experience heightened levels of discomfort; only participants rated as introverts did. The findings of this study suggested that individual characteristics can have a great influence on strategies for dissonance reduction; people who are more extraverted are less susceptible to feelings of cognitive dissonance.

Measures of extraversion. Self-report measures assess extraversion. Matz et al. (2008) used a commonly used version of the Big Five Inventory (John & Srivastava, 1999). This measure lists 44 characteristics, and people complete the measure by indicating on a 5-point Likert scale the extent to which they agree that each characteristic describes them. There are 8 items that comprise the extraversion dimension: "is talkative," "is reserved," "is full of energy," "generates a lot of enthusiasm," "tends to be quiet," "has an assertive personality," "is sometimes shy, inhibited," and "is outgoing, sociable." Since the results of the Martz et al. study suggested that individuals who are more talkative, outgoing, energetic, enthusiastic, assertive, uninhibited, and sociable may be less prone to experience cognitive dissonance, extraverted SPs who use projective instruments when they are aware that these test do not have sound psychometric properties may feel less discomfort in doing this than their more introverted colleagues.

Self-Efficacy

Self-efficacy theory. Another possible explanation for SPs' use of projective instruments is that they may be convinced that they are able to use projective tests to make desirable recommendations for students. Bandura (1977) defined self-efficacy as "the conviction that one can successfully execute the behavior required to produce the outcomes" (Bandura, 1977, p. 193). Further:

> Self-efficacy refers to beliefs in one's capabilities to mobilize the motivation, cognitive resources, and courses of action needed to meet given situational demands. Self-beliefs of efficacy affect the challenges that are undertaken, the amount of effort expended in an endeavor, the level of perseverance in the face of difficulties, whether thinking patterns take self-aiding or self-impeding forms, and vulnerability to stress and depression. (Bandura & Wood, 1989, p. 408)

In still other words, "self-efficacy refers to beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 3).

Bandura (1997) delineated the multiple dimensions that efficacy beliefs can have. The first dimension is *level*, which refers to the idea that "the range of perceived capability for a given person is measured against levels of task demands that represent varying degrees of challenge or impediment to successful performance" (p. 42). The second dimension is *generality*, which refers to the idea that "people may judge themselves efficacious across a wide range of activities or only in certain domains of functioning" (p. 43). The last dimension is *strength*, which refers to the idea that "weak efficacy beliefs are easily negated by disconfirming experiences, whereas people who have a tenacious belief in their capabilities will persevere in their efforts despite innumerable difficulties and obstacles" (Bandura, 1997, p. 43). Chen, Gully, and Eden (2001) summarized these three dimensions as: the level or magnitude of the difficulty of the task at hand, the strength of one's certainty of being able to perform the task successfully, and the generalizability of the magnitude and strength dimensions across other situations.

Bandura (1997) also described the four sources of self-efficacy beliefs. First, people can build self-efficacy through enactive mastery experiences, which provide people with firsthand evidence of whether they can succeed at a given task. In enactive mastery experiences, "successes build a robust belief in one's personal efficacy" but "failure undermines it, especially if failures occur before a sense of efficacy is firmly established" (Bandura, 1997, p. 80). Second, people can build self-efficacy through vicarious experiences, which provide people with models of others performing a certain task. In vicarious experiences, "people must appraise their capabilities in relation to the attainments of others" (Bandura, 1997, p. 84). Third, people can build self-efficacy through verbal persuasion, in which others provide support and bolster one's faith in one's self. In verbal persuasion, self-efficacy increases "if significant others express faith in one's capabilities" (Bandura, 1997, p. 101). Fourth, and last, self-efficacy relies on physiological and affective states; how people read their physiological and affective reactions can determine their feelings of self-efficacy. That is, "because high arousal can debilitate performance, people are more inclined to expect success when they are not beset by aversive arousal than if they are tense and viscerally agitated" (Bandura, 1997, p. 106).

In line with Bandura's theory, people will choose to engage in activities at which they expect to be successful, and they will avoid situations where they believe they cannot be successful. Further, once they choose to engage in an activity, people's feelings of self-efficacy help determine how active they are and how hard they try to be successful (Bandura, 1997). In addition, self-efficacy can transfer from one situation to another;

bolstering feelings of mastery in one situation can affect a person's behavior in another, dissimilar situation (Bandura, 1977).

Self-efficacy research. The concept of self-efficacy may be beneficial in studying why SPs do not follow practice guidelines. That is, if practitioners do not believe they will be successful at a certain practice, then they will not engage in it. Conversely, if SPs believe they can interpret projective measures to make useful educational recommendations, then they are likely to use those measures. Unfortunately, there is a dearth of peer-reviewed literature as this concept relates to SPs; however, a recent doctoral dissertation (Huber, 2006) examined self-efficacy and control beliefs of SPs. Using a sample of 297 graduate students and working professionals, Huber (2006) collected data using the Huber Inventory of Self-Efficacy for School Psychologists (HIS-SP) – a measure she created as part of her dissertation – and the Spheres of Control Scale-Version 3 (SOC-3). The HIS-SP measured self-efficacy of SPs in particular by asking about the specific functions that SPs undertake, and the SOC-3 measured personal efficacy, interpersonal control, and sociopolitical control. Huber (2006) reported alpha reliabilities of .71, .77, and .61 for the three scales, respectively, and test-retest correlations above .90 for a 4-week interval. For the HIS-SP, Huber (2006) found it had a clear factor structure and internal consistency above .90. The results of this dissertation showed higher feelings of self-efficacy for practicing SPs than for graduate students, but no differences in feelings of control. More importantly, however, the study introduced a new measure of self-efficacy for use specifically with SPs, which could be of great use in this dissertation.

Despite the lack of research about self-efficacy as it relates to SPs' adherence to practice guidelines, investigators have studied this phenomenon in other fields. For example, Cabana et al. (1999) reviewed published research that examined barriers to physician adherence to practice guidelines. They organized the barriers by effect on knowledge, attitudes, and behavior. The effect on knowledge included lack of familiarity and lack of awareness. That is, physicians may not be aware of appropriate guidelines and even if they are aware of them, they may not be familiar with how to apply them correctly. The effect on attitudes included lack of agreement with specific guidelines or guidelines in general, lack of outcome expectancy, lack of self-efficacy, and lack of motivation/inertia of previous practice. Physicians did not necessarily agree with a certain guideline, or with guidelines in general – in the studies the authors reviewed, lack of agreement with a specific guideline was more common than lack of agreement with the idea of guidelines in general. The effect on behavior included external barriers: guideline-related barriers, patient-related barriers, and environmental-related barriers. Physicians may have encountered situations in their day-to-day practice that prevented them from adhering to guidelines (Cabana, 1999).

According to Cabana et al. (1999), self-efficacy "is the belief that one can actually perform a behavior. It influences whether a behavior will be initiated and sustained despite poor outcomes" (p. 1462). In their review, Cabana and colleagues found that physicians were less likely to follow a guideline if they thought it would not lead to a better outcome. In addition, physicians with low self-efficacy may adhere less to guidelines, and physicians' low self-efficacy because of low confidence in their ability or preparation may cause lower adherence.

Measures of self-efficacy. Self-report instruments measure self-efficacy. One such scale is the Self-Efficacy Scale, which Sherer et al. (1982) developed. This scale factors self-efficacy into two subscales: general self-efficacy and social self-efficacy. Participants answer items on a 14-point Likert scale. In their first study of the measure, the authors found Cronbach alpha reliability coefficients of .86 for the general scale and .71 for the social scale. To determine the scale's construct validity, the authors also administered other scales measuring locus of control, social desirability, and self-esteem, among others. Results suggested that none of the scales correlated enough to determine that they were measuring the same constructs, which is the result the authors hoped for to demonstrate construct validity. In a second study, the authors examined the criterion validity of the scale; they found higher scores on the General Self-Efficacy Scale correlated with a higher chance of being employed, and a lower chance of quitting or being fired from jobs.

Chen, Gully, and Eden (2001) evaluated Sherer et al.'s scale and revised it to devise the New General Self-Efficacy Scale (NGSES). The NGSES looks at self-efficacy as a trait, not a state quality (i.e., self-efficacy is not situation-specific in this study of the development of the NGSES). General self-efficacy, however, is thought to predict specific efficacy. The authors noted that, "accumulation of successes in life, as well as persistent positive vicarious experiences, verbal persuasion, and psychological states, augment" (p. 63) feelings of general self-efficacy, which are pervasive and affect peoples' feelings about their potential for personal mastery across domains.

On the NGSES, participants answer eight items on a 5-point Likert scale with responses ranging from strongly disagree (1) to strongly agree (5). Total scores for the

scale can range from 8 to 40 points. In one study of the measure, the authors found single-factor solution alphas from .85 to .88 and test-retest reliability coefficients from .62 to .65. Content validity ratings resulted from students' judgments about what the items were measuring: 98% were self-efficacy, 2% were self-esteem. In a second study of their measure, the authors found internal consistency reliability coefficients of .86 and .90, at two different test times. They also found a test-retest coefficient of .67. In general, the authors' revision of the Self-Efficacy Scale resulted in a unidimensional, internally consistent, and stable measure of general self-efficacy.

As mentioned above, Huber (2006) developed a self-efficacy scale specifically for SPs. On the HIS-SP (Huber, 2006) participants answer 75 items on a 7-point Likert scale with responses ranging from not well at all (1) to very well (7). Huber (2006) constructed the questions "based upon the major roles and functions of school psychologists obtained from the literature through a deductive approach" (p. 53). Factor-analysis showed that there were five factors in the scale: Intervention and Consultation Skills, Multidimensional Assessment Skills, Counseling Skills, Professional Interpersonal Skills, and Research Skills. Huber (2006) reported coefficient alphas for these factors as .96, .94, .91, .93, and .90, respectively, and .98 for the total test. Huber (2006) provided a solid measure with which to examine SPs' self-efficacy feelings, particularly in the realm of assessment.

Dissertation Pilot

In a preliminary study, Whitney and Tryon (2009) conducted a pilot study to examine the assessment practices of SPs. Participants in the study were 27 practicing SPs in New York State whose names and email addresses were available publicly on school

district websites. I contacted 100 potential participants by email; the email contained a description of the study and of the participants' rights. The email also contained a link to a questionnaire on Surveymonkey.com. Of the 100 participants I recruited, 28 responded by beginning the questionnaire. One person did not fully complete the questionnaire, leaving a final sample of 27.

The questionnaire asked participants to rate cognitive and social-emotional assessment measures with regard to frequency of use and usefulness. Participants also indicated how important they thought various aspects of tests to be (e.g., ease of interpretation of results). The questionnaire further included demographic characteristics, graduate training (university, degree, and theoretical orientation), employment setting, and years in practice.

The majority of the respondents were women (n = 21, 78%) and were Caucasian (n = 25, 93%). Most of the participants held masters or specialist degrees (n = 22, 82%), and all had degrees involving the study of school psychology (School Psychology: n = 24; School-Clinical-Child Psychology: n = 3).

The results of this pilot study showed that SPs reported using projective measures despite their lack of psychometric evidence. Two participants reported using the TAT or CAT "Often," 2 participants reported using the HTP "Often," and 10 participants reported using the Sentence Completion "Often." In addition, in general, people were most confident about the properties of the instruments that they use. The correlation between use and self-efficacy for some objective measures, for example the WAIS-III, was positive and significant, $r(25) = .44$, $p = 0.024$, as it was for some projective measures, for example the HTP, $r(25) = 0.46$, $p = 0.017$. In addition, it seems that SPs

who use the projective instruments, particularly the Rorschach, more frequently had less concern with social-emotional measures having strong psychometric properties, $r(25) = -0.396$, $p = 0.041$, good validity, $r(25) = -0.456$, $p = 0.017$, or an appropriate norming sample, $r(25) = -0.499$, $p = 0.008$.

The results also showed that SPs' theoretical orientation was not significantly related to their use of projective tests. Using ANOVA, however, I found a significant difference between the judgments of a measure's psychometric validity importance depending on an SP's theoretical orientation, $F(1, 4) = 4.003$, $p = 0.014$. Using Tukey post-hoc comparisons, I found the significant differences were between cognitive-behavioral and psychoanalytic orientations (Difference =1.47, std.error = 0.409, $p = 0.013$), and between systems-focused and psychoanalytic orientations (Difference = 1.67, std.error = 0.507, $p = 0.025$).

Finally, the results showed that SPs have different reasons for choosing cognitive measures and social-emotional measures (i.e., SPs will rely more on judgment issues for projective test selection and more on psychometric information for objective instrument selection). Using paired comparisons, I found that school psychologists rated cognitive measures as more important in having strong psychometric properties, $t(26) = 2.267$, $p = 0.032$, good test-retest reliability, $t(26) = 2.842$, $p = 0.009$, and being required in the workplace, $t(26) = 2.078$, $p = 0.048$.

The findings of this study showed that SPs continue to use various projective measures to assess social-emotional functioning. In addition, SPs tended to have different standards for the psychometric properties of objective and projective assessment tools, or cognitive and social-emotional ones, and for the reasons that they choose said measures.

Lastly, the results show that the theoretical orientation of one's training program affects one's views on important test characteristics. Although this study extended the literature on assessment practices of SPs, it did not adequately explore why SPs continue to use measures for which there is little or no psychometric validation. In addition, the sample was small, and only comprised SPs practicing in New York.

Rationale and Hypotheses

There is a considerable amount of research that examines the assessment practices of SPs; however, thus far, no studies have examined why SPs continue to use projective tests in schools when there is overwhelming evidence against their use in the school setting combined with practice guidelines and cautions against their use. It is especially puzzling to consider why SPs still use projective measures when more objective, accurate, research-based measures exist (Kamphaus & Frick, 2005; Lachar & LaCombe, 1983; Reynolds & Kamphaus, 2004). Levine (2003) proposed that some psychiatrists, psychologists, and other mental health professionals engage in self-deception "to espouse attitudes, behave therapeutically, or promulgate expert opinions in ways that are less than justified or not substantiated by empirical research" (p. 325), but he does not cite or provide any research support. Wenger and Pryzwansky (1987) surveyed SPs about their implementation of the APA Specialty Guidelines for the Delivery of Services by School Psychologists, but this study did not address assessment in particular. As such, it is timely and important to examine the assessment practices of SPs and the reasons that could explain why they utilize projective measures. Consequently, the current study attempted to fill the gap in the literature. I hypothesized that:

H01: Some SPs will report that they use projective measures as part of school assessments.

H02: SPs who spent more time on campus during graduate school will use projective tests less often than those who spent less time on campus during graduate school.

H03: SPs' use of projective measures will be negatively correlated with their feelings of cognitive dissonance about using projective measures.

H04: SPs' feelings of self-efficacy, both general and SP-specific will be positively correlated with their use of projective measures.

H05: SPs' self-ratings of extraversion will be negatively correlated with reported feelings of cognitive dissonance about using projective measures.

H06: SPs will make trade-offs between psychometric properties and clinical judgment in their choices between hypothetical tests.

Chapter 3: Method

This chapter presents the methodology that this study used to examine the relationship between usage of social-emotional assessment measures, cognitive dissonance, extraversion, and self-efficacy amongst SPs. The chapter begins with a description of the selection of participants and characteristics of the respondents, including demographic variables and professional identity. This is followed by a description of the instrument, including the development and description of the social-emotional hypothetical test usage measures, and the study procedures. The chapter concludes with a section on data analysis.

Participant Selection and Characteristics of Respondents

After receiving approval from the Institutional Review Board of the City University of New York Graduate School and University Center, I solicited participation from practicing SPs. I sent an email (see Appendix A) to presenters whose emails were listed in the program for the National Association of School Psychologists (NASP) 2010 Annual Convention. This email contained a description of the study, the participants' rights, and a link to the questionnaire on Surveymonkey.com.

To ensure confidentiality, participants did not provide any identifying information. Participants who had questions about the study or wanted to be informed of the results could contact me directly by email without having to link their questions to their responses, thus ensuring their privacy and confidentiality of responses.

I needed at least 100 participants to detect a medium effect size at the $p < .01$ level of significance (Cohen, 1992). I sent the email to 915 people, of which 120 (13%) completed the questionnaire. There were four respondents who completed about half of

the questionnaire but then stopped. Had they finished the questionnaire, the final sample would have totaled 124. Because there was no identifiable reason for why these people did not complete the questionnaire, and because there were only four of them, statistical comparison of this subset with the completers would not be meaningful. In addition, there were four participants who indicated that they held a bachelors degree as their highest degree; because one needs at least a masters degree to be a practicing school psychologist, I deleted these four participants from the data analysis. This left a final sample of 116 (13%).

Demographic information. Participants completed questions pertaining to their personal, educational, and professional backgrounds. Where available, I also included the data from recent NASP Membership Surveys in these tables in order to determine if my sample was characteristic of NASP members. Table 1 shows the personal variables including age, sex, ethnicity, and geographic region of residence. The majority of SPs who completed the entire survey was Caucasian, female, and between the ages of 20 and 39. A plurality resided in the Southeast. Relative to NASP members, the sample had a higher percentage of women and Southeastern residents, and a lower percentage of Western residents.

Table 1

Personal Demographic Characteristics of Questionnaire Completers

Demographic	Variable	*N*	%	NASP Membership Data[a]
Age	20-29	39	33.62	Mean = 46.2
	30-39	50	43.10	
	40-49	15	12.93	
	50-59	8	6.90	
	60-69	4	3.45	
Gender	Male	18	15.52	26.00
	Female	98	84.48	74.00
Ethnicity	Caucasian	96	82.76	92.6
	Asian	4	3.45	0.9
	African-American	12	10.34	1.9
	Hispanic	2	1.72	3.0
	Other	2	1.72	0.8
Region of Residence	Central	30	25.86	21.79
	Northeast	32	27.59	32.89
	Southeast	44	37.93	21.72
	West	10	8.62	22.09

Note: N = 116.
[a]Data for Region of Residence from Fagan & Wise (2007). All other NASP data from Curtis et al. (2008).

Table 2 shows the educational demographic variables, including highest degree earned, geographic region where that degree was earned, type of training program, and theoretical orientation of training program. A plurality of SPs who completed the entire

survey had doctoral degrees and attended graduate school in the Southeast. The majority was trained in School Psychology in graduate programs with either a cognitive-behavioral or systems focus. Relative to NASP members, sample members tended to have a higher percentage of doctoral degrees.

Table 2

Educational Demographic Characteristics of Questionnaire Completers

Demographic	Variable	N	%	NASP Membership Data[a]
Highest Degree Earned	Masters	31	26.72	32.6
	Specialist	27	23.28	34.9
	Doctorate	58	50.00	32.4
Training Program	School Psychology	109	93.97	N/A
	School-Clinical Child Psychology	2	1.72	N/A
	Other	5	4.31	N/A
Training Orientation	Cognitive-Behavioral	36	31.03	N/A
	Psychoanalytic	1	0.86	N/A
	Systems-Focused	35	30.17	N/A
	Eclectic	21	18.10	N/A
	Behavioral	13	11.21	N/A
	Other	10	8.62	N/A
Region Where Highest Degree Was Earned	Central	34	29.31	N/A
	Northeast	36	31.03	N/A
	Southeast	41	35.34	N/A
	West	5	4.31	N/A

Note: N = 116.

[a]NASP member data from Curtis et al. (2008).

Table 3 shows the professional demographic variables, including primary workplace, geographic region of primary workplace, years working as a SP, ranking of time spent in various activities, and participation in and reasons for professional development. A plurality of SPs who completed the survey worked in elementary schools and worked in the Southeast. Their years working as SPs ranged from 0 to 31 (M = 6.86, SD = 7.14). Curtis et al. (2008) reported that the mean length of experience for their sample from the NASP Membership Survey was 14.8 years. Professional activities that consumed most of their professional time were: Assessment/Testing, which respondents most frequently ranked first, Consultation, which respondents most frequently ranked second, Counseling, which respondents most frequently ranked third, and Crisis Intervention, which respondents most frequently ranked fourth. Almost all respondents (96%) also reported that they participated in continuing education (CE), and 56% reported they did so "often." The most commonly endorsed reason for participating in CE was to keep up with current intervention practices.

Table 3

Professional Demographic Characteristics of Questionnaire Completers

Demographic	Variable	*N*	%
Primary Workplace	Preschool	6	5.17
	Elementary School	47	40.52
	Middle School	8	6.90
	High School	12	10.34
	Hospital	3	2.59
	Private Practice	2	1.72
	Other	38	32.76

Table 3 (continued)

Demographic	Variable	*N*	%
Region of Primary Workplace	Central	30	25.86
	Northeast	31	26.72
	Southeast	45	38.79
	West	10	8.62
Ranking of Time Spent in Major Activities	Assessment/Testing: Rank 1	59	50.86
	Consultation: Rank 2	54	46.55
	Counseling: Rank 3	55	47.41
	Crisis Intervention: Rank 4	66	56.90
Participates in CE	Yes	113	95.69
	No	5	4.31
Frequency of CE Participation	Never	-	-
	Rarely	6	5.17
	Sometimes	45	38.79
	Often	65	56.03
Reason for CE Participation*	Required by Law	57	49.14
	Required by Employer	41	35.34
	Keep Up with Current Intervention Practices	99	85.34
	Keep Up with Current Assessment Practices	88	75.86

	Range	M	SD
Years Working as a SP	0-31	6.86	7.14

Note: N = 116. *Percentages add to more than 100 because respondents could choose more than one option.

Professional identity. To assess participants' sense of professional identity, the questionnaire contained questions about membership in NASP and the American Psychological Association (APA). This section also asked about what credentials the participants possessed, that is, a certificate and/or a license. Professional organization membership and credentials are both indicators of professional identity (VanZandt, 1990). Because graduate school years are formative for professional identity (Bruss & Kopala, 1993; Colbeck, 2008), this section also asked how many hours a day and days a week participants' spent on campus during graduate school. Last, participants answered questions about how much they thought graduate school contributed to their professional identity development and how much they thought experiences since leaving graduate school have contributed to their professional identity development. Table 4 shows that most responders were members of NASP, but most were not members of APA. The majority held a state certificate, but not a national certificate or state license. Most completers also indicated their professional identity depended a lot on experiences both in graduate school and after. The hours a day that completers reported spending on campus during graduate school ranged from 1 to 20 ($M = 6.94$, $SD = 3.55$) and the days a week ranged from 1 to 7 ($M = 4.24$, $SD = 1.13$); I deleted 8 values above 24 and 4 values above 7 for these variables, respectively. Relative to NASP members, a higher percentage of the sample had state certification and licensure.

Table 4

Professional Identity Variables

Variable	Response	*N*	%	NASP Membership Data[a]
Member of APA	Yes	41	35.34	N/A
	No	75	64.66	N/A
Member of NASP	Yes	101	87.07	100.00
	No	15	12.93	0.00
Holds State Certificate	Yes	81	69.83	91.30
	No	35	30.17	8.70
Holds National Certificate	Yes	48	41.38	N/A
	No	68	58.62	N/A
Holds State License	Yes	34	29.31	36.70
	No	82	70.69	63.30
Professional Identity as a Result of Graduate School Training and Experience	Not at all	-	-	N/A
	Very little	4	3.45	N/A
	Somewhat	50	43.10	N/A
	A lot	62	53.45	N/A

Table 4 (continued)

Variable	Response	*N*	%	
Professional Identity as a Result of Experiences Since Leaving Graduate School	Not at all	3	2.59	N/A
	Very little	7	6.03	N/A
	Somewhat	46	39.66	N/A
	A lot	60	51.72	N/A
	Range	*M*	*SD*	
Hours a Day Spent on Campus During Graduate School (*N* = 108)	1-20	6.94	3.55	N/A
Days a Week Spent on Campus During Graduate School (*N* = 112)	1-7	4.24	1.13	N/A

Note: N = 116.

[a]NASP member data from Curtis et al. (2008).

Instrument

An online research questionnaire that I developed with Dr. Tryon and used in my pilot study was the primary data collection measure (see Appendix B). I modified the questionnaire to include changes suggested at the proposal for this dissertation. The research questionnaire included eight measures: (a) demographic information (given above in Participants section), (b) professional identity information (also presented in Participants section), (c) social-emotional test preferences, (d) social-emotional test usage, (e) cognitive dissonance, (f) extraversion and other personality characteristics, (g) general self-efficacy, and (h) SP-specific self-efficacy.

Social-emotional test preferences. This portion of the questionnaire helped to reveal respondents' preferences by forcing them to choose between hypothetical tests. This design, known as conjoint measurement, is a form of stated choice method that historically has been used in the marketing field to determine consumers' preferences for products (Green & Srinivasan, 1978, 1990; Green & Wind, 1975). The method forces respondents to make trade-offs that can "provide…valuable information about the relative importance of various attributes of a product" and "can also provide information about the value of various levels of a single attribute" (Green & Wind, 1975, p. 108). This is a useful tool in fields other than marketing (Louviere, 2000), and it was of particular relevance in this study because it helped to reveal the preferences of the study's participants. By having to make trade-offs between the various levels of the three attributes, the respondents revealed the factors that are most important to them in choosing social-emotional tests.

Development of the social-emotional test preferences measure. This part of the questionnaire examined what test attributes were more important to participants by requiring them to make trade-offs among attributes when choosing an instrument for use. Participants read descriptions of hypothetical tests that varied systematically along three attributes: (a) psychometric properties, (b) convenience, and (c) degree of clinical judgment. Participants then chose the test that they would prefer to use in practice (see Appendix B, pp. 120-132).

The hypothetical tests had (a) two levels of psychometric properties: excellent or adequate; (b) two levels of how convenient they were: very or moderately; and (c) three levels of required clinical judgment: a lot, a moderate amount, or very little (see

Appendix C). I will provide a more detailed description of each of these levels in the following section.

Design of the pair-wise comparisons resulted in 12 total hypothetical tests, which are listed in Appendix D. If all possible comparisons of the 12 tests were used, there would have been 66 pair-wise comparisons, many more than the maximum of 15 pair-wise comparisons recommended in the literature regarding this type of research design (Kuhfeld, 2005). As such, working with Dr. Verkuilen, I decided to have 12 versions of the questionnaire with at least 11 but no more than 13 of the pairs in each version. To determine which pairs to include in each of the first 6 forms of the questionnaire, we first listed all 66 pairs and numbered the first 6 from 1 to 6. For each following group of 6 pairs, we reordered the numbering to minimize the correlations between the combinations of 1 to 6. We used this method to ensure that each hypothetical test appeared the fewest number of times within each questionnaire. For forms 7 through 12 of the questionnaire, we counterbalanced the items from each of the first 6 forms to ensure that each pair-wise comparison appeared roughly the same amount of times across the questionnaires. Specifically, form 7 consisted of the odd numbered pairs from form 1 and the even numbered pairs from form 4; form 8 consisted of the odd numbered pairs from form 2 and the even numbered pairs from form 5; form 9 consisted of the odd numbered pairs from form 3 and the even numbered pairs from form 6; form 10 consisted of the even numbered pairs from form 1 and the odd numbered pairs from form 6; form 11 consisted of the even numbered pairs from form 2 and the odd numbered pairs from form 5; and form 12 consisted of the even numbered pairs from form 3 and the odd numbered pairs from form 4.

We also decided on two pairs that we wanted to have in each questionnaire because of their importance in this study. Based on Dr. Tryon's clinical experience, we chose these comparisons because she believed that practitioners might endorse some trade-offs in reliability and validity in order to be able to exercise some clinical judgment. The first pair of items included a test with excellent psychometric properties that is very convenient and requires very little clinical judgment compared with a test that has adequate psychometric properties, is very convenient, and requires moderate clinical judgment. The second pair of items included a test with excellent psychometric properties that is very convenient and requires moderate clinical judgment compared with a test that has adequate psychometric properties, is very convenient, and requires a lot of clinical judgment. Appendix E presents the resulting distribution of the pair-wise comparisons to questionnaires. Questionnaires 1 through 5, and 7, 8, 11, and 12 have 13 pair-wise comparisons each; Questionnaire 9 has 12 pair-wise comparisons; and Questionnaires 6 and 10 each has 11 pair-wise comparisons.

Description of the hypothetical social-emotional tests. Table 5 shows the definitions I used in creating the hypothetical social-emotional tests (see also Appendix C). The table shows the three different attributes and the corresponding descriptions for each level of each attribute.

Table 5
Characteristics of Hypothetical Tests

Attribute	Level	Description
Psychometric Properties	Excellent	The test has reliability coefficients of .90, correlates highly with measures of similar constructs, and has norms based on a sample very similar to the student you are testing.
	Adequate	The test has reliability coefficients of .80, correlates moderately with measures of similar constructs, and has norms that reflect the general school population.
Convenience	Very	You have 100 copies on hand in your office, and it takes 10 minutes to administer and score.
	Moderate	You purchase test protocols for $25 per a pack of 5, and it takes 20 minutes to administer and score.
Clinical Judgment	A lot	Items are ambiguous and allow students to respond freely. The manual provides suggestions but permits you to exercise your judgment in administration and scoring.
	Moderate	Items are standard and allow students to respond freely. The manual provides a standard administration protocol and suggestions for scoring that require you to exercise judgment in assigning scores.
	Very little	Items are standard and students respond by marking one of several alternatives. The manual provides exact instructions for standard administration and scoring.

Last, in this part of the questionnaire, respondents answered three questions indicating which test attributes were most important to them (see Appendix B, pp. 133-135). They responded to the questions on a 5-point scale, with 1 corresponding to totally preferring one attribute, and 5 corresponding to completely preferring the other attribute. They indicated their preference between psychometric properties and clinical judgment

(p. 133), between psychometric properties and convenience (p.134), and between clinical judgment and convenience (p.135).

Social-emotional test usage. To determine participants' actual use of particular assessments and the importance they placed on the properties of these assessments, the next section asked respondents if they knew of specific social-emotional assessment measures (see Appendix B, p. 136) and whether they had used those measures in the past year (see Appendix B, p. 137). Respondents also ranked their top three choices of tests with respect to how confident they were in their abilities to administer and interpret the results of the measures (see Appendix B, p. 138), how good they perceived the psychometric properties of the measures to be (see Appendix B, p. 139), and which measures they found most useful (see Appendix B., p. 140). Last, participants ranked the top three choices to specify what they thought were the most important features of the specified tests (e.g., has an appropriate norming sample, has good validity, etc.) when choosing a social-emotional assessment measure (see Appendix B, p. 141).

The projective measures included in this part of the questionnaire were based on similar studies that examined projective test usage (Camara, Nathan, & Puente, 2000; Cashel, 2002; Goh, Teslow, & Fuller, 1981; Hojnoski, 2006; Prout, 1983; Sweeney, Clarkin, & Fitzgibbon, 1987; Watkins, Campbell, Nieberding, & Hallmark, 1995; Wilson & Reschly, 1996; Vukovich, 1983) and studies that have reviewed the psychometric properties of commonly used projective measures (Halperin & McKay, 1998; Lilienfeld et al., 2000; Sattler & Hoge, 2006). The list of tests also included two objective measures, the BASC-2 and the Achenbach Child Behavior Checklist, both of which have

established psychometric properties and norming populations, in order to allow comparisons between the use of projective measures and the use of objective measures.

Cognitive dissonance. Because people often experience cognitive dissonance when they behave in ways that conflict with their cognitions (Festinger & Carlsmith, 1959), I measured cognitive dissonance using a portion of an unnamed emotion measure (Elliott & Devine, 1994; Matz, Hofstedt, & Wood, 2008) that assesses negative self-evaluation, positive emotions, and dissonance discomfort (see Appendix B, pp. 142-145). Participants answered five items that assessed how uneasy, uncomfortable, tense, bothered, and concerned they felt on a 7-point Likert scale with responses ranging from 1 (*does not apply at all*) to 7 (*applies very much*). Total scores for the scale range from 1 to 35, and the authors reported a coefficient alpha of .81. The alpha for the scale in this study was .94. The following is an example of this question in this questionnaire:

> For this question, please imagine that you are required to administer and interpret a test that does not have norms that are representative of the population you are using the test for. Please mark the extent to which each of the words below applies to how you are feeling currently. (See Appendix B, p. 142).

Matz et al. (2008) used these types of questions to measure how participants were feeling in the moment. For the purposes of this study, I used these questions to assess how people felt when asked to imagine themselves administering and interpreting tests that require considerable clinical judgment, do not have standardized administration or scoring procedures, do not have adequate norms, or do not have adequate psychometric evidence. These four features are salient features of projective testing techniques (Kamphaus & Frick, 2005; Lilienfeld et al., 2000).

Extraversion. I measured extraversion with a frequently used measure of the Big Five Personality Traits (John & Srivistava, 1999; Matz et al., 2008). Though Matz et al. (2008) found that only extraversion was related to feelings of cognitive dissonance, I included the entire measure for this study (see Appendix B, pp. 146-147). First, this helped to disguise the fact that I was interested in studying extraversion, and second, this allowed me to explore possible relationships of other personality traits with projective test usage. In addition to extraversion, this part of the questionnaire assessed agreeableness, conscientiousness, neuroticism, and openness.

Participants answered 44 items on a 5-point Likert scale with responses ranging from 1 (*disagree strongly*) to 5 (*to agree strongly*). Total scores for the extraversion scale range from 8 to 40 points, and the authors report a coefficient alpha of .89. The alpha in this study also was .89 for the extraversion scale. The alphas for the other scales were as follows: .80 for agreeableness, .82 for openness, .83 for neuroticism, and .83 for conscientiousness.

Self-efficacy. Sherer et al. (1982) developed the Self-Efficacy Scale, which factored into two subscales: general self-efficacy and social self-efficacy. In their first study of the measure, the authors found Cronbach alpha reliability coefficients of .86 for the general scale and .71 for the social scale.

Chen, Gully, and Eden (2001) revised Sherer et al.'s scale and evaluated it to devise the New General Self-Efficacy Scale (NGSES; see Appendix B, p. 148). The NGSES looks at self-efficacy as a trait, not a state quality. Participants answer eight items on a 5-point Likert scale with responses ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Total scores for the scale can range from 8 to 40 points. In one study of the

measure, the authors found single-factor solution alphas from .85 to .88 and test-retest reliability coefficients from .62 to .65. Content validity ratings had students judge what the items were measuring: 98% were self-efficacy, 2% were self-esteem. Coefficient alpha for this scale in this study was .90.

Huber (2006) developed a self-efficacy scale specific to SPs, and found it had a clear factor structure and internal consistency above .90. Relevant to this study, I included the 18 items of this scale that measure Multidimensional Assessment Skills (see Appendix B, p. 149); these items measure "the complex process of assessment involving administration and scoring of assessment measures, as well as, encompassing ethical and legal guidelines, interpretation skills, report writing, and assessment related decision-making" (Huber, 2006, p. 107). Coefficient alpha for this portion of the scale in this study was .94.

Procedure

As the principal investigator for this study, I sought approval from the Institutional Review Board of the City University of New York Graduate School and University Center. After I obtained this approval, I sought participation from practicing SPs and collected data by emailing SPs whose emails were available in the 2010 NASP Convention brochure. After entering the email addresses into a spreadsheet in alphabetical order by last name, I assigned each address a number from 1 to 12, corresponding to the form of the questionnaire I sent to each person. I started by assigning a 1 to the first person on the list, and counted up to 12 for the following names on the list. Each time I reached the number 12, I started the numbering again.

Participants completed the research questionnaire online. The first page of the questionnaire explained the study and the requirements of participation, and completion of the questionnaire was considered each participant's informed consent. The first page of the questionnaire also indicated that I was the principal investigator for the study, but that I would not have access to any identifying information from the participants. Participants did not provide any identifying information by completing the questionnaire, so there was no need to code the data for confidentiality. I designed the survey so that participants could not go back to change their answers to previous questions that they had already answered.

The questionnaire took approximately 20 minutes to complete. Participants who completed the survey and chose to enter their email addresses at the end were entered into a drawing to receive one of three $25 Amazon.com gift certificates. I provided a website for respondents to copy and paste into their browsers where they could enter their email addresses. This ensured that the participants' email addresses were detached from their answers in order to preserve their confidentiality. Of the 116 who completed the questionnaire, 83 (72%) entered their email addresses. I downloaded the responses of the completed questionnaires into a spreadsheet, which I then transferred to SAS to analyze the responses.

Data Analysis

I used various statistical methods to test the hypotheses in this study. First, I provided descriptive statistics for the demographic variables including age, sex, ethnicity, state of residence, highest degree earned, state where degree was earned, type of training program, theoretical orientation of training program, primary workplace, state of primary

workplace, years working as a SP, percent of time spent in various activities, and participation in and reasons for professional development. Second, I provided descriptive statistics for participants' rankings of the various cognitive and social-emotional measures. Third, I provided descriptive statistics and intercorrelations for participants' responses to the questions pertaining to cognitive dissonance, self-efficacy, and personality factors.

The major analysis of this dissertation used logistic regression. By gathering information from respondents on their preferences between hypothetical social-emotional tests, I was able to perform conjoint measurement. Conjoint measurement forces respondents to make trade-offs between the various levels of the different attributes, thereby revealing their preferences (Green & Wind, 1975; Green & Srinivasan, 1978; Green & Srinivasan, 1990). Logistic regression is the method by which to study conjoint measurement. In this model, participants' choices between the hypothetical tests with different characteristics were the dependent variables, and the hypothetical tests themselves were the independent variables.

Chapter 4: Results

The primary aim of this study was to examine the reasons why SPs continue to use projective tests. Another goal of this study was to examine SPs' trade-offs between test characteristics. This chapter provides descriptive statistics and results for the hypotheses of this study.

Descriptive Statistics of Questionnaire Completion

Table 6 shows how many people completed each questionnaire. Questionnaire number 7 had considerably more responses than the other questionnaires. Because I sent each of the 12 questionnaires to approximately the same number of people, and because most of the people who completed questionnaire 7 reported living in Georgia, one possible explanation for this phenomenon is that a respondent or respondents to questionnaire 7 may have forwarded the email containing the survey link to colleagues.

Table 6

Breakdown of Questionnaires Completed

Questionnaire #	N	%
1	13	11.21
2	14	12.07
3	2	1.72
4	8	6.90
5	7	6.03
6	5	4.31
7	36	31.03
8	7	6.03
9	5	4.31
10	6	5.17
11	8	6.90
12	5	4.31

Note: $N = 116$.

Descriptive Statistics of Questions Related to Projective Tests

Table 7 shows that most respondents rated psychometric properties as more important than both convenience and use of clinical judgment, and clinical judgment as more important than convenience.

Table 7

Preference of Test Attributes

Comparison	Range	*M*	*SD*
Psychometric Properties (1) v. Convenience (5)	1-4	1.82	0.71
Psychometric Properties (1) v. Clinical Judgment (5)	1-4	2.19	0.81
Convenience (1) v. Clinical Judgment (5)	1-5	3.27	0.99

Note: N = 116.

Table 8 shows that most respondents had heard of the measures presented, and the BASC-2 was the most commonly used measure in the 2009-2010 school year.

Table 8

Knowledge and Use of Common Social-Emotional Measures

Knowledge/Use	Test	*N*	%	% Missing
Heard of:	BASC-2	114	98.28	0.86
	Rorschach	112	96.55	0.86
	TAT	109	93.97	0.86
	CAT	98	84.48	1.72
	Bender-Gestalt	113	97.41	0.00
	Kinetic Family Drawing	100	86.21	0.86
	HTP	110	94.83	0.86
	Draw a Person	112	96.55	0.86
	Achenbach	115	99.14	0.00
	Sentence Completion	110	94.83	0.00
	RAT	87	75.00	1.72
Used in the 09-10 school year:	BASC-2	102	87.93	0.00
	Rorschach	4	3.45	10.34
	TAT	9	7.76	8.62
	CAT	5	4.31	10.34
	Bender-Gestalt	41	35.34	6.03
	Kinetic Family Drawing	31	26.72	7.76
	HTP	32	27.59	7.76
	Draw a Person	33	28.45	6.03
	Achenbach	44	37.93	4.31
	Sentence Completion	52	44.83	4.31
	RAT	18	15.52	9.48

Table 9 shows that the BASC-2 received the most number 1 rankings, and the Achenbach garnered the most number 2 rankings for most comfort administering and scoring, best psychometric properties, and most useful. Sentence completion received the most number 3 rankings for comfort administering and scoring, while the Bender-Gestalt received the most number 3 rankings for best psychometric properties and being most useful.

Table 9

Participants' Most Frequent Rankings of Social-Emotional Test Characteristics

Characteristic	Test and Rank	*N*	%
Comfort Administering and Scoring	BASC-2: Rank 1	91	78.44
	Achenbach: Rank 2	47	40.52
	Sentence Completion: Rank 3	24	20.69
Psychometric Properties	BASC-2: Rank 1	88	75.86
	Achenbach: Rank 2	74	63.79
	Bender-Gestalt: Rank 3	61	52.59
Most Useful	BASC-2: Rank 1	90	77.59
	Achenbach: Rank 2	63	54.31
	Bender-Gestalt: Rank 3	36	31.03

Table 10 shows that participants most frequently ranked “addresses referral question” as the first most important social-emotional test characteristic, “strong psychometric properties” as the second most important characteristic, and “appropriate norming sample” as the third most important characteristic.

Table 10

Participants’ Rankings of Top 3 Most Important Social-Emotional Test Characteristics

Rank	Characteristic	*N*	%
1	Addresses Referral Question	54	46.55
2	Strong Psychometric Properties	32	27.59
3	Appropriate Norming Sample	31	26.72

Descriptive Statistics of Cognitive Dissonance, Personality, and Self-Efficacy Variables

Table 11 shows that the cognitive dissonance that participants experienced from imagining administering test with inadequate norms, inadequate psychometric evidence, and no standardized administration or scoring procedures yielded the highest mean scores. The cognitive dissonance experienced from imagining administering a test requiring considerable clinical judgment yielded the lowest mean score. Higher scores on this measure indicate more feelings of cognitive dissonance.

Table 11

Cognitive Dissonance Variables

Cognitive Dissonance Experienced From Test that Have:	Range	*M*	*SD*
Inadequate Norms	5-35	28.48	6.67
Inadequate Psychometric Evidence	7-35	28.33	5.82
No Standardized Administration or Scoring Procedures	5-35	25.90	7.34
Considerable Clinical Judgment	1-35	19.41	8.67
Total Cognitive Dissonance	20-140	102.12	22.40

Note: $N = 116$.

Table 12 shows the results of the personality measure, where higher scores correspond to a participant rating him or herself as having more of that scale's characteristics. Participants rated themselves highest on characteristics of conscientiousness. Next highest were ratings of agreeableness, then openness, and then extraversion. Participants rated themselves lowest on characteristics of neuroticism. Based on a median split of the extraversion score (Matz, Hofstedt, & Wood, 2008) 51 of

respondents qualified as extraverts and 65 qualified as introverts. Applying this median split to the other personality variables resulted in the following: 57 were above the median and 59 were below for agreeableness, exactly half of the respondents fell above and below the median for conscientiousness, 57 were above the median and 59 were below for neuroticism, and 55 were above the median and 61 were below for openness.

Table 12

Extraversion and Other Personality Variables

Variable	Range	*Median*	*M*	*SD*
Extraversion	11-40	27.00	27.35	6.86
Agreeableness	24-45	38.00	37.15	4.90
Conscientiousness	24-45	39.50	38.31	5.20
Neuroticism	8-37	19.00	20.16	6.01
Openness	23-50	36.00	36.77	5.97

Note: $N = 116$.

Table 13 shows respondents' ratings of general self-efficacy and SP self-efficacy, where higher scores correspond to a participants' greater feelings of self-efficacy. Because the highest possible scores were 40 and 90 for general and SP self-efficacy, respectively, this shows that most participants rated themselves high on feelings of both general and SP self-efficacy.

Table 13

Self-Efficacy Variables

Variable	Range	*M*	*SD*
General Self-Efficacy	24-40	36.47	3.68
SP Self-Efficacy	54-90	82.86	6.51

Note: $N = 116$.

Correlations Among Study Variables

Table 14 shows the correlations among cognitive dissonance, self-efficacy, and extraversion. The correlation between SP self-efficacy and general self-efficacy was positive and significant, indicating that practitioners who possessed greater feelings of self-efficacy were likely to feel that way in general and about specific SP skills. In addition, the correlations between extraversion and both general and SP self-efficacy were positive and significant, indicating that respondents who rated themselves higher on the extraversion scale also possessed greater feelings of self-efficacy.

Table 14

Correlations Among Cognitive Dissonance, Self-Efficacy, and Extraversion

	Total Cognitive Dissonance	General Self-Efficacy	SP Self-Efficacy	Extraversion
Total Cognitive Dissonance	1			
General Self-Efficacy	0.12	1		
SP Self-Efficacy	-0.02	0.50***	1	
Extraversion	0.03	0.25**	0.20*	1

Note: $N = 116$.

* $p < .05$. ** $p < .01$. *** $p < .0001$.

Table 15 shows the correlations among the cognitive dissonance variables. The positive and significant correlations between all the variables suggest that respondents who experienced cognitive dissonance from one of the factors listed also experienced cognitive dissonance from the other factors.

Table 15

Correlations Among Cognitive Dissonance Variables

	Inadequate Norms	Inadequate Psychometric Evidence	No Standardized Administration or Scoring Procedures	Considerable Clinical Judgment
Inadequate Norms	1			
Inadequate Psychometric Evidence	0.74**	1		
No Standardized Administration or Scoring Procedures	0.53**	0.57**	1	
Considerable Clinical Judgment	0.35*	0.36**	0.47**	1

Note: $N = 116$.

* $p < .001$. ** $p < .0001$.

Table 16 shows the correlations among the personality variables. These results show that people who rated themselves as high in the extraversion category also rated themselves as being more agreeable and open, but less neurotic. In addition, people who rated themselves as being agreeable also rated themselves as being more conscientious and less neurotic. Lastly, people who rated themselves as being more conscientious also rated themselves as being less neurotic, and people who rated themselves as being more open rated themselves as being less neurotic.

Table 16

Correlations Among Personality Variables

	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness
Extraversion	1				
Agreeableness	0.22*	1			
Conscientiousness	0.00	0.25**	1		
Neuroticism	-0.37***	-0.36***	-0.23*	1	
Openness	0.25**	0.10	0.06	-0.19*	1

Note: $N = 116$.

* $p < .05$. ** $p < .01$. *** $p < .0001$.

Analyses Related to Study's Hypotheses

Table 17 shows the results of correlations (corrected for attenuation) that tested H02, the hypothesis that people who spent more time on campus would utilize projective tests less. Respondents who reported spending more days a week on campus during graduate school were less likely to have used the HTP and the RAT during the 2009-2010 school year.

Table 17
Correlations between Time Spent on Campus During Graduate School and Use of Projective Tests in 09-10 School Year

	Hours a Day	Days a Week
Rorschach	-0.08 ($N = 97$)	-0.14 ($N = 100$)
TAT	-0.29 ($N = 99$)	-0.35 ($N = 102$)
CAT	0.24 ($N = 97$)	0.06 ($N = 100$)
Bender-Gestalt	-0.03 ($N = 102$)	-0.15 ($N = 105$)
Kinetic Family Drawing	0.01 ($N = 100$)	-0.09 ($N = 103$)
HTP	-0.23 ($N = 100$)	-0.35* ($N = 103$)
Draw a Person	-0.03 ($N = 102$)	-0.07 ($N = 105$)
Sentence Completion	-0.15 ($N = 104$)	-0.17 ($N = 107$)
RAT	-0.07 ($N = 98$)	-0.44* ($N = 101$)

* $p < .01$.

Table 18 shows the results of correlations (corrected for attenuation) that tested H03, the hypothesis that SPs' use of projective measures would be negatively correlated with their feelings of cognitive dissonance about using projective measures. This table shows that people who administered the Bender-Gestalt, HTP, Draw a Person, and

Sentence Completion experienced less cognitive dissonance from imagining themselves administering a test that required considerable clinical judgment. In addition, people who administered the Draw a Person experienced less total cognitive dissonance than other people.

Table 18
Correlations between Cognitive Dissonance Variables and Use of Projective Tests in 09-10 School Year

	Inadequate Norms	Inadequate Psychometric Evidence	No Standardized Administration or Scoring Procedures	Considerable Clinical Judgment	Total
Rorschach ($N = 106$)	0.00	0.12	0.25	-0.45	-0.07
TAT ($N = 104$)	-0.03	-0.07	-0.14	-0.31	-0.19
CAT ($N = 106$)	0.18	0.19	0.38	-0.10	0.17
Bender-Gestalt ($N = 109$)	-0.13	-0.10	-0.06	-0.35**	-0.21
Kinetic Family Drawing ($N = 107$)	-0.05	0.01	-0.13	-0.22	-0.15
HTP ($N = 107$)	0.22	0.07	-0.16	-0.31*	-0.10
Draw a Person ($N = 109$)	-0.08	-0.13	-0.24*	-0.36**	-0.28*
Sentence Completion ($N = 111$)	-0.07	-0.07	-0.17	-0.32**	-0.22
RAT ($N = 105$)	-0.14	-0.01	-0.09	-0.30*	-0.18

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

Table 19 shows the results of correlations (corrected for attenuation) that tested H04, the hypothesis that SPs' use of projective tests would be positively correlated with their feelings of self-efficacy. The table shows that only respondents who administered the HTP in the 2009-2010 school year had higher feelings of SP self-efficacy.

Table 19

Correlations between Self-Efficacy and Use of Projective Tests in 09-10 School Year

	General Self-Efficacy	SP Self-Efficacy
Rorschach (N = 104)	-0.02	-0.15
TAT (N = 106)	-0.05	0.32
CAT (N = 104)	0.37	0.31
Bender-Gestalt (N = 109)	0.11	0.17
Kinetic Family Drawing (N = 107)	0.20	0.16
HTP (N = 107)	0.23	0.31*
Draw a Person (N = 109)	0.13	0.25
Sentence Completion (N = 111)	0.16	0.17
RAT (N = 105)	0.09	0.28

* $p < .05$.

Table 20 shows the results of the correlations that tested H05, the hypothesis that SPs' self-ratings of extraversion would be negatively correlated with reported feelings of cognitive dissonance. There were no significant relationships between the extraversion variable and the cognitive dissonance variables.

Table 20

Correlations between Cognitive Dissonance and Extraversion

Dissonance Concerning	Extraversion
Inadequate Norms	0.06
Inadequate Psychometric Evidence	0.05
No Standardized Administration or Scoring Procedures	-0.06
Considerable Clinical Judgment	0.05

Note: $N = 116$.

This next section shows the analyses used to explore H06, the hypothesis that SPs will make trade-offs between psychometric properties and clinical judgment in their choices between hypothetical tests. I used logistic regressions to analyze the paired comparisons that respondents made. In this model of paired comparisons, respondents expressed their values of objects by choosing between the pair; the choice was coded as a binomial variable where a value of 1 indicated the respondent's preference (Anderson, 2009). First, I performed a baseline regression in which test choice was the dependent variable and the actual tests themselves were the independent variables. For each observation, the pair of tests being compared was coded with a 1 or -1, and the choice was coded 1 if the respondent chose the test with value 1 or 0 if the respondent chose the test with value -1. I dropped Test 10 from the regression because it was the worst test

(adequate psychometric properties, moderately convenient, requires a lot of clinical judgment).

Table 21 shows the results of this baseline regression. The coefficients illustrate that respondents preferred every other test to Test 10. Further, the comparisons of the coefficients show that, all other factors held constant, people preferred tests with better psychometric properties. Examining the following test pairs shows these comparisons: 1 and 7, 2 and 8, 3 and 9, 4 and 10, 5 and 11, and 6 and 12. In addition, this table shows that, all other factors held constant, people preferred tests that were more convenient. Examining the following test pairs shows this result: 1 and 4, 2 and 5, 3 and 6, 7 and 10, 8 and 11, and 9 and 12. This table also shows that, all other factors held constant, people preferred tests that required less clinical judgment. Examining the following sets of tests shows this result: 1, 2, and 3; 4, 5, and 6; 7, 8, and 9; and 10, 11, and 12.

Lastly, examining the pairs that were included in all the questionnaires (2 and 7; 3 and 8) reveals that respondents preferred the tests that required less clinical judgment. They were not willing to compromise psychometric properties to be able to exercise some clinical judgment.

Table 21

Baseline Model Parameter Estimates and Standard Errors with Consensus Rankings

Parameter	Test Description	Estimate	Standard Error	Consensus Ranking
Test 3	P+, C+, J-	5.297	0.364	1
Test 2	P+, C+, Jo	4.628	0.343	2
Test 6	P+, C-, J-	4.313	0.352	3
Test 5	P+, C-, Jo	3.620	0.339	4
Test 9	P-, C+, J-	3.541	0.352	5
Test 1	P+, C+, J+	2.934	0.333	6
Test 8	P-, C+, Jo	2.855	0.332	7
Test 4	P+, C-, J+	2.838	0.340	8
Test 12	P-, C-, J-	2.765	0.338	9
Test 11	P-, C-, Jo	2.087	0.313	10
Test 7	P-, C+, J+	1.508	0.337	11
Test 10	P-, C-, J+	0.00	-	12

Model Fit Statistics	
Degrees of Freedom	11
AIC	1232.725
-2log Likelihood	1210.725

Note: P = Psychometric properties, with + indicating excellent, and - indicating adequate.

C = Convenience, with + indicating very and - indicating moderately.

J = Clinical judgment, with - indicating very little, o indicating moderate, and + indicating a lot.

Next I performed the same analysis using the following groupings: geographic region of residence, highest degree earned, and theoretical orientation of training program. I created the geographic region variable based on Fagan and Wise's (2007) reporting of the 2006 NASP membership data, in which the states belong to four regions: Northeast, Southeast, Central, and West. For highest degree earned, I combined all the doctoral degrees because there were not enough observations in the Psy.D. and Ed.D. categories. For theoretical orientation of training programs, I combined the one psychoanalytic person with the "other" category, and I combined cognitive-behavioral with behavioral because the latter did not have enough data in it to be its own category.

Table 22 shows how respondents in each of the geographic regions ranked the various tests. Respondents in the Central, Northeast, and West geographic regions always preferred tests that required less clinical judgment, holding the other factors constant. Respondents in the Southeast, however, sometimes preferred tests that required more clinical judgment.

Table 22

Ranking of Tests by Geographic Region

Test Rank	Central ($N = 30$)	Northeast ($N = 32$)	Southeast ($N = 44$)	West ($N = 10$)
1	P+, C+, J-	P+, C+, J-	P+, C+, Jo	P+, C+, J-
2	P+, C-, J-	P-, C+, J-	P+, C+, J-	P+, C-, J-
3	P+, C+, Jo	P+, C+, Jo	P+, C-, J-	P+, C+, Jo
4	P+, C-, Jo	P+, C-, J-	P+, C-, J+	P-, C+, J-
5	P-, C-, J-	P-, C+, Jo	P+, C-, Jo	P+, C-, Jo
6	P-, C+, J-	P+, C-, Jo	P-, C+, J-	P+, C+, J+
7	P+, C+, J+	P-, C-, J-	P+, C+, J+	P-, C+, Jo
8	P-, C+, Jo	P+, C-, J+	P-, C+, Jo	P-, C-, J-
9	P+, C-, J+	P+, C+, J+	P-, C-, J-	P-, C-, Jo
10	P-, C-, Jo	P-, C-, Jo	P-, C-, Jo	P+, C-, J+
11	P-, C+, J+	P-, C+, J+	P-, C+, J+	P-, C+, J+
12	P-, C-, J+	P-, C-, J+	P-, C-, J+	P-, C-, J+

Note: P = Psychometric properties, with + indicating excellent, and - indicating adequate.

C = Convenience, with + indicating very and - indicating moderately.

J = Clinical judgment, with - indicating very little, o indicating moderate, and + indicating a lot.

Table 23 shows the correlations among the rankings of the tests by geographic region. As seen in the table, there was a statistically significant correlation between the rankings in the Southeast and Central regions.

Table 23

Correlations Among Rankings by Geographic Region

	Central	Northeast	Southeast	West
Central	1			
Northeast	0.30	1		
Southeast	0.45*	-0.12	1	
West	0.00	0.03	0.12	1

Note: $N = 12$.

* $p < .05$.

Table 24 shows how respondents with different graduate degrees ranked the various tests. Respondents in all three groups of graduate degrees always preferred tests that required less clinical judgment, holding the other factors constant.

Table 24

Ranking of Tests by Highest Degree Earned

Test Rank	Doctorate ($N = 58$)	Masters ($N = 31$)	Specialists ($N = 27$)
1	P+, C+, J-	P+, C+, J-	P+, C+, J-
2	P+, C+, Jo	P+, C+, Jo	P+, C-, J-
3	P+, C-, J-	P+, C-, J-	P+, C-, Jo
4	P+, C-, Jo	P-, C+, J-	P-, C+, J-
5	P-, C+, J-	P+, C-, Jo	P+, C+, Jo
6	P-, C+, Jo	P+, C+, J+	P+, C-, J+
7	P+, C+, J+	P-, C-, J-	P-, C+, Jo
8	P-, C-, J-	P-, C+, Jo	P+, C+, J+
9	P+, C-, J+	P+, C-, J+	P-, C-, J-
10	P-, C-, Jo	P-, C-, Jo	P-, C-, Jo
11	P-, C+, J+	P-, C+, J+	P-, C+, J+
12	P-, C-, J+	P-, C-, J+	P-, C-, J+

Note: P = Psychometric properties, with + indicating excellent, and - indicating adequate.

C = Convenience, with + indicating very and - indicating moderately.

J = Clinical judgment, with - indicating very little, o indicating moderate, and + indicating a lot.

Table 25 shows the correlations among the rankings of the tests by highest degree earned. As seen in the table, there were no significant correlations between the groups.

Table 25

Correlations Among Rankings by Highest Degree Earned

	Doctorate	Masters	Specialist
Doctorate	1		
Masters	0.24	1	
Specialist	-0.15	0.30	1

Note: N = 12.

Table 26 shows how respondents with different theoretical training ranked the various tests. Respondents in the behavioral/cognitive-behavioral, eclectic and other categories always preferred tests that required less clinical judgment, holding the other factors constant. Respondents with systems-focused training, however, sometimes preferred tests that required more clinical judgment.

Table 26

Ranking of Tests by Theoretical Orientation of Training Program

Test Rank	Behavioral/Cognitive-Behavioral ($N = 49$)	Eclectic ($N = 21$)	Systems-Focused ($N = 35$)	Other ($N = 11$)
1	P+, C+, J-	P+, C+, J-	P+, C+, Jo	P+, C+, J-
2	P+, C+, Jo	P+, C+, Jo	P+, C+, J-	P+, C-, J-
3	P+, C-, J-	P+, C-, J-	P+, C-, J-	P-, C+, J-
4	P+, C-, Jo	P+, C-, Jo	P-, C+, J-	P+, C+, Jo
5	P-, C+, J-	P-, C+, J-	P+, C-, Jo	P+, C-, Jo
6	P+, C+, J+	P+, C+, J+	P-, C+, Jo	P+, C+, J+
7	P-, C+, Jo	P+, C-, J+	P-, C-, J-	P-, C+, Jo
8	P+, C-, J+	P-, C-, J-	P+, C-, J+	P+, C-, J+
9	P-, C-, J-	P-, C+, Jo	P-, C-, Jo	P-, C-, J-
10	P-, C-, Jo	P-, C-, Jo	P+, C+, J+	P-, C-, Jo
11	P-, C+, J+	P-, C+, J+	P-, C+, J+	P-, C+, J+
12	P-, C-, J+	P-, C-, J+	P-, C-, J+	P-, C-, J+

Note: P = Psychometric properties, with + indicating excellent, and - indicating adequate.

C = Convenience, with + indicating very and - indicating moderately.

J = Clinical judgment, with - indicating very little, o indicating moderate, and + indicating a lot.

Table 27 shows the correlations among the rankings of the tests by theoretical orientation of training group. The correlations between behavioral/cognitive-behavioral and eclectic, and between behavioral/cognitive-behavioral and other were positive and significant.

Table 27

Correlations Among Rankings by Theoretical Orientation of Training Program

	Behavioral / Cognitive-Behavioral	Eclectic	Systems-Focused	Other
Behavioral / Cognitive-Behavioral	1			
Eclectic	0.58*	1		
Systems-Focused	0.24	-0.12	1	
Other	0.67*	0.36	0.15	1

Note: $N = 12$.

* $p < .01$.

Additional Analyses

To further explore the differences between respondents who administered which tests in the 2009-2010 school year, I compared the groups of people who administered these tests according to certain variables. Because of the frequency of use of the various tests, I was only able to include certain tests in these analyses; some tests had either two many or too few people who indicated they had used the test, hence making any analysis using these data not meaningful. The tests I included in these analyses were the Draw a Person, Kinetic Family Drawing, Bender-Gestalt, House-Tree-Person, Sentence Completion, and Achenbach. The variables on which I compared the usage of these tests were the following: hours a day spent on campus during graduate school, days a week

spent on campus during graduate school, professional identity as a result of graduate school, professional identity as a result of post-graduate school, total cognitive dissonance, general self-efficacy, SP self-efficacy, and extraversion. Table 28 shows which variables were significantly different between the groups who used or did not use these specific tests (see Appendix F for the complete results of these analyses). Respondents who administered the Draw a Person had lower total cognitive dissonance scores than those who did not. Further, respondents who administered the House-Tree-Person spent fewer days a week on campus during graduate school, felt less strongly that their professional identity was a result of graduate school, and had higher feelings of SP self-efficacy.

Table 28

Significant Differences Between Variables According to Test Usage

Test	Comparison Variable	Difference Between Groups	*t* Value	Significance Level
Draw a Person ($N = 109$)	Total Cognitive Dissonance	9.49	2.16	< .05
House-Tree-Person ($N = 103$)	Days a Week on Campus	0.58	2.51	< .05
House-Tree-Person ($N = 107$)	Professional Identity as a Result of Graduate School	0.32	2.76	< .01
House-Tree-Person ($N = 107$)	SP Self-Efficacy	-2.71	-2.12	< .05

I also created examined the characteristics of respondents based on how many projective tests they reported using in the 2009-2010 school year. To do this, I created a variable that summed the number of tests each respondent reported using, and that variable was the dependent variable in a Poisson zero-inflated regression. (I also looked

at a more detailed model that used a latent variable to make the scale rather than a sum, and the traits essentially were the same). Table 29 shows the results of this analysis. People who administered more projective tests experienced more cognitive dissonance from imagining themselves using a test that required considerable clinical judgment.

Table 29

Zero-Inflated Poisson Regression Estimates and Standard Errors

Parameter	Estimate	Standard Error
Intercept	1.51	4.12
Hours a Day Spent on Campus During Graduate School	0.02	0.07
Days a Week Spent on Campus During Graduate School	0.21	0.22
Cognitive Dissonance from Inadequate Norms	-0.03	0.06
Cognitive Dissonance from Inadequate Psychometric Evidence	-0.01	0.07
Cognitive Dissonance from No Standardized Administration or Scoring Procedures	0.02	0.04
Cognitive Dissonance from Considerable Clinical Judgment	0.08*	0.03
Extraversion	-0.03	0.04
General Self-Efficacy	-0.11	0.08
SP Self-Efficacy	0.01	0.05

Model Fit Statistics	
Degrees of Freedom	80
AIC	366.85
-2log Likelihood	326.85

* $p < .05$.

Last, I looked at the comment field at the end of the survey and classified respondents' statements. Seven (6.03%) respondents entered text into this field. Three of

the comments were respondents' concerns about the wording of some of the questionnaire items, three respondents provided information about the types of assessments they conduct and measures they prefer to use, and one respondent wished me luck.

Summary of Findings Related to Study's Hypotheses

Table 30 presents this study's hypotheses and indicates that one hypothesis was supported, three were partially supported, and two were not supported. I tested H01 with a frequency count of respondents who administered projective tests during the 2009-2010 school year. The frequencies of people who administered the Rorschach, TAT, CAT, Bender-Gestalt, Kinetic Family Drawing, HTP, Draw a Person, Sentence Completion, and RAT were greater than one, supporting this hypothesis.

I tested H02 using correlations. I found a correlation of $r = -0.24$ significant at the $p < .05$ level between usage of the HTP and days a week spent on campus during graduate school, and correlation of $r = -0.27$ significant at the $p < .05$ level between the usage of RAT and days a week spent on campus during graduate school. This partially supported H02.

I tested H03 also using correlations. I found significant correlations between cognitive dissonance from having to use considerable clinical judgment and usage of the Bender-Gestalt ($r = -0.27$, $p < .01$), HTP ($r = -0.22$, $p < 0.5$), Draw a Person ($r = -0.26$, $p < 0.1$), Sentence Completion ($r = -0.26$, $p < 0.1$), and RAT ($r = -0.20$, $p < 0.5$). In addition, there was a correlation of $r = -0.20$ significant at the $p < 0.5$ level between total cognitive dissonance and usage of the Draw a Person. This partially supported this hypothesis.

I tested H04 using correlations as well. I found a correlation of $r = 0.20$ significant at the $p < 0.5$ level between usage of the HTP and SP self-efficacy, partially supporting this hypothesis.

I tested H05 using correlations. There were no significant correlations in this analysis; hence, this hypothesis was not supported.

I tested H06 using conjoint measurement and logistic regression. I found that, on the whole, respondents preferred tests that required less clinical judgment, suggesting that this hypothesis was not supported.

Table 30
Overview of Study Hypotheses

H0 Number	Study Hypothesis	Supported/ Not Supported
H01	Some SPs will report that they use projective measures as part of school assessments.	Supported
H02	SPs' who spent more time on campus during graduate school will use projective tests less than those who spent less time on campus during graduate school.	Partially Supported
H03	SPs' use of projective measures will be negatively correlated with their feelings of cognitive dissonance about using projective measures.	Partially Supported
H04	SPs' feelings of self-efficacy, both general and SP-specific, will be positively correlated with their use of projective measures.	Partially Supported
H05	SPs' self-ratings of extraversion will be negatively correlated with reported feelings of cognitive dissonance about using projective measures.	Not Supported
H06	SPs will make trade-offs between psychometric properties and clinical judgment in their choices between hypothetical tests.	Not Supported

Chapter 5: Discussion

In this chapter I will present the key findings resulting from the statistical analyses in the present study. In addition, I will present the implications of the findings, the limitations of the study, and directions for future research.

Key Findings

The purpose of this study was to examine the reasons why SPs continue to use projective tests despite limited evidence of their reliability and validity (Kamphaus & Frick, 2005; Lilienfeld et al., 2000; Sattler & Hoge, 2006). In particular, I wanted to explore how professional identity, cognitive dissonance, self-efficacy, and extraversion related to the test choices that practitioners made. Another goal of this study was to examine SPs' trade-offs between test characteristics, and how the same variables (professional identity, cognitive dissonance, self-efficacy, and extraversion) related to these trade-offs.

This sample appeared to be a conscientious group of professionals who, for the most part, followed the practice guidelines and laws governing their profession. The majority reported participating often in continuing education because they wanted to keep up with current assessment and intervention practices. Most also were members of NASP, the governing organization of the profession. Though the latter was a function of the way I collected the sample (i.e., soliciting participants from NASP convention presenters), these factors nonetheless indicate that the sample as a whole was concerned with current practices in the field.

In general, when presented with hypothetical test choices, the population sampled behaved in a way that aligned with ethical and professional expectations. That is, when

given the choice between hypothetical tests, the majority of respondents chose the norm-referenced, psychometrically more robust tests. Nonetheless, some practitioners did report continued use of actual projective tests, with the largest number endorsing use of Sentence Completion, Bender-Gestalt, Draw a Person, House-Tree-Person (HTP), and Kinetic Family Drawing.

The majority of respondents in this study had heard of all the social-emotional tests presented to them, including both projective and objective tests. In terms of usage, almost everyone reported administering the BASC-2 during the 2009-2010 school year. Further, most people rated the BASC-2 as the easiest to score and administer, the most useful, and having the best psychometric properties. In addition, respondents indicated that it was most important to them that a social-emotional test address their referral question, have strong psychometric properties, and have an appropriate norming sample.

As noted above, however, many people also reported administering some of the projective tests during the 2009-2010 school year. In some cases, indicators of professional identity and cognitive dissonance were negatively correlated with projective test usage, and indicators of self-efficacy were positively correlated with projective test usage. The House-Tree-Person (HTP) garnered the most statistically significant results in this regard (the RAT had similar effect sizes across many of the variables, but the results were not always statistically significant). People who reported using the HTP spent fewer days a week on campus during graduate school, felt their professional identity was less a result of graduate school, had higher feelings of SP self-efficacy, and reported lower feelings of cognitive dissonance from using a test that required considerable clinical judgment such as the HTP.

It is worthwhile to consider why the HTP, and sometimes the RAT, were the few projective tests that related to the other variables I measured in this study. It is possible that because so few people used some of the other projectives, like the Rorschach, there were not enough data to find a relationship. It is also possible that the HTP is one of the least psychometrically sound, but convenient, projectives of all the tests presented in this study. The test simply requires a person to draw a house, a tree, and a person; the evaluator then interprets the responses. It is possible that this level of convenience draws SPs into using the test despite its lacking any reliability, validity, or norms. This could be particularly true if practitioners are using the HTP as an icebreaker when working with younger children; having a child draw can be a very useful way of building rapport and starting a session in a non-threatening way.

Overall, respondents reported a preference for the hypothetical test that had the best psychometric properties, was the most convenient to administer and score, and required the least amount of clinical judgment. Conversely, respondents consistently ranked last the hypothetical test with the worst psychometric properties, the least convenience, and the most required amount of clinical judgment. In addition, it did not appear that respondents were willing to make trade-offs in psychometric properties to be able to exercise some clinical judgment in the administration and interpretation of a test. When broken down by geographic region and theoretical orientation of training program, there were a few instances when the preferences did not align with this pattern; that is, sometimes people preferred a test requiring more clinical judgment holding the other factors constant. This, however, was the exception, and for the most part, the preferences were consistent in favor of psychometrically sound instruments.

Support for the hypotheses and theoretical underpinnings of this study was mixed. Research has reported that SPs spend most of their professional hours performing assessments (Bramlett et al., 2002; Koonce, 2007; Fagan & Wise, 2007; Reschly, 2000), and that certainly was the case with participants in this study. In addition, research has shown that SPs utilize projective measures in assessments (Hojnoski, Morrison, Brown, & Matthews, 2006; Shapiro & Heick, 2004), and this study also confirmed this phenomenon.

What was not as clear, however, was why SPs continue to use projective measures, which was the most important question that led me to do this study. Despite rating strong psychometric properties as a priority for social-emotional test characteristics, some respondents still reported using tests that lack these characteristics. In some cases, SPs who felt more affiliated with graduate school, both by spending more time on campus and feeling that it formed their professional identity (Bruss & Kopala, 1993; Colbeck, 2008) reported using projective measures less.

Cognitive dissonance theory also explained some of the respondents' behavior, in that people often experience cognitive dissonance when they behave in ways that conflict with their cognitions (Festinger & Carlsmith, 1959). This was true in this study in some cases; cognitive dissonance as a result of using a test that required considerable clinical judgment was negatively associated with use of that test. This makes sense in that respondents who experience greater cognitive dissonance from using projective tests most likely would not be the ones reporting that they actually used those tests. Conversely, those who reported using projective tests had less cognitive dissonance concerning this practice.

Research on personality traits has shown that extraversion often moderates the effects of cognitive dissonance (Matz et al., 2008). This generally did not prove to be the case in this study. In addition, self-efficacy theory and research (Cabana et al., 1999; Huber, 2006) suggest that people will choose to engage in activities at which they expect to be successful (Bandura, 1997). Again, the relationship between self-efficacy and projective test usage in this study was limited. It is possible, however, that this is a result of respondents' generally high ratings of themselves on feelings of both general and SP self-efficacy; it is possible that there was not enough variation in these reported scores to find a relationship between the factors.

Perhaps one of the study's greatest strengths was that it provided a different way of examining assessment practices that is more similar to how people actually practice. The paired comparison research design (Green & Srinivasan, 1978, 1990; Green & Wind, 1975; Louviere, 2000), using conjoint measurement for data analysis **(**Anderson, 2009; Kuhfeld, 2005**)**, has never been used before to study SPs' testing preferences. Because practitioners make trade-offs every day in their practice and choices between tests, this design is a more relevant and robust method of examining testing practices than simply asking people which tests they use. This is a design that should be used by people who study how people make choices. It is a more relevant design to practitioners because they make trade-offs every day in their practice.

Implications for School Psychology

SPs spend the majority of their time performing assessments that can have a substantial effect on a child's educational trajectory. As such, it was important to understand SPs' practices and preferences in choosing assessment instruments. These

practices and preferences are especially important to understand when they seemingly contradict the best practices, laws, and guidelines that govern the profession. Though studies exist in other professions that examine why practitioners do not adhere to practice guidelines, no study has ever examined this phenomenon amongst SPs.

As a practicing SP myself, I was eager to find out more about why SPs continue to use projective tests, and what variables might explain some SPs' self-confidence in using and interpreting measures that have no standardized procedures for doing so. Any insight that this study could provide into why SPs use projective tests would benefit all SPs, practicing or not, because it would help the field to understand what is preventing people from adhering to the guidelines and laws that have been set out. The information this study yielded concerning the relationship amongst projective test usage, professional identity, cognitive dissonance, extraversion, and self-efficacy could help in educating SPs so that they are better able to follow the laws and guidelines. The significant relationships amongst these variables found in this study also could help training programs identify people who might be prone to using the projective tests, and provide an opportunity for training programs to design targeted education for those people about the reasons why projective tests do not fit the practice guidelines and laws.

Results from this study suggest that creating cognitive dissonance in students about use of projective tests might be an effective deterrent to their usage. This might be done by continually emphasizing the lack of validity evidence for projective tests. In addition, there was some relationship between time spent on campus and a decreased use of projective tests. This suggests that training programs could target students who spend less time than others on campus during graduate school, and encourage them to become

more involved in departmental activities with both peers and faculty. The students who spend more time on campus arguably will be altered more by the graduate school experience, and therefore end up with views and practices that align with the university's orientation. Reaching the students who might already have formulated their own views, as indicated by their lack of affiliation within their graduate schools, could be one way that the field of SP could use the information from this study to inform its practice. Training professionals could initiate discussions with these students concerning their lack of involvement and perhaps require student attendance at departmental activities.

Limitations of the Current Study

There were several limitations to this study. One major limitation is that the sample only included people who presented at the 2010 NASP conference, and hence may have reflected only a certain type of conscientious professional who is tuned into practice guidelines and standards for practice. This could have led to the limited variation evident in many of the item responses, for example, usage of certain projective tests. This in turn limited the number of analyses I could perform to detect differences between groups of people who used projective tests and those who did not.

A second limitation related to the sample is that there may have been a social desirability bias in many of the respondents' answers. Because many people associated with NASP know what they are *supposed* to do, it is possible that their answers to many of the questions, especially the paired comparisons of the hypothetical tests, reflected this knowledge rather than what they *actually* do.

A third limitation of this study is that the questionnaire did not explicitly ask why SPs chose to use certain projective measures; that is, there was nowhere for SPs to

indicate their reasons for utilizing these tests in the 2009-2010 school year. As such, there was limited information with which to draw conclusions about why respondents chose to use certain projective tests but not others.

A fourth limitation of this study is that each respondent only got a subset of the test comparisons. It was not feasible to ask respondents to make 66 hypothetical paired test comparisons for risk of losing participants partway through the questionnaire. As a result, however, there were not enough data points for each of the pairs to thoroughly analyze the data in the desired way.

A fifth limitation of this study is related to the software I used to collect the data. The website, surveymonkey, was immensely helpful in many ways, but lacked certain features that would have been ideal. For example, there was no way to prevent people from forwarding the email containing the survey link to colleagues. This may explain why one survey had considerably more respondents than the others.

Directions for Future Research

Suggestions for future research address the identified limitations of and expand the findings from this study. First, future research on this topic should include a more diverse sample of practicing SPs, and not just those that present at conferences for professional organizations like NASP. This would yield a more diverse sample set in terms of SPs' practices, and it would allow for more extensive analysis of the resulting data. This kind of sample might be difficult to garner, but the resulting data could have much more variation in terms of reported behavior and preferences, hence making some of the analyses more fruitful.

Second, if possible, in future research, respondents should receive all of the hypothetical paired comparisons so the resulting data is complete enough to perform more in-depth analyses. This might mean dropping one of the levels of the clinical judgment variable in order to have fewer total pairs. This loss of a level of this variable could be worth it, however, if the resulting dataset allows for a more complete conjoint measurement analysis.

Third, future instruments should incorporate a measure of social desirability, a measure of how frequently respondents use projective tests, and explicit questions that measure the reasons SPs choose to use the projective tests they do. The latter could include alternatives that respondents could mark in response to reasons for using each test. It could also draw on the writing of Kleinmutz (1990), who proposes various reasons for why people continue to use their "heads instead of formulas" (p. 302). Future research could assess his stated reasons, which include deluded self-confidence, the complexities of decisions, and the costs of assessing decision-making.

Fourth, other software options might be worth exploring to determine if their functionality is better than the software I used in this study. In addition, it is worth exploring if it is possible to design a randomization process for assigning respondents to questionnaires. This might involve directing all respondents to the same webpage, which in turn randomly assigns each respondent to one of the questionnaires.

Fifth, future researchers would benefit from using, where possible, inventories with published norms. This would allow researchers to get a better sense of how the respondents rate themselves on cognitive dissonance, self-efficacy, and personality characteristics.

Sixth, a future study of this topic might want to utilize a different method of presenting the hypothetical tests. This study used a non-branded way of forced comparisons (Green &Wind, 1975); the hypothetical tests did not have identifiable names attached to them. Further, respondents did not make their choices based on scenarios in which they would be using the tests. The study used this design because there would have been a voluminous amount of text for respondents to read if they had to read both vignettes and test descriptions for each paired-comparison item. A future study, however, could utilize vignettes if the study used a branded method of forced comparisons. In that study, respondents could read scenarios and then choose between actual, named tests. Such a design would be able to explore whether practitioners choose different tests based on different circumstances.

Conclusion

This study sought to further knowledge of SPs' assessment practices, particularly with regard to social-emotional test usage. In particular, the study examined what factors relate to SPs' use of social-emotional tests and what trade-offs practitioners were willing to make between psychometric properties, convenience, and clinical judgment in their choices of hypothetical social-emotional tests.

Overall, the study demonstrated that when faced with hypothetical test choices, most respondents behaved in ways that indicated the sample was one of knowledgeable professionals who recognize what is expected of them in terms of testing practices. Nonetheless, although SPs preferred tests to have robust psychometric properties, they continued to use projective tests that, by nature, are lacking in this respect. The constructs of professional identity, cognitive dissonance, self-efficacy, and extraversion, and the

relationships among those variables and between those and the usage of projective tests provided some explanation for this phenomenon.

Lastly, this study provided an innovative method for examining the choices that SPs make in their everyday practices. The measurement of trade-offs is a technique that should be used in future studies examining these phenomena.

Appendix A
Introductory Email

My name is Sarah Whitney, and I am a student in the School Psychology Ph.D. Program in the Educational Psychology Department at The Graduate Center of the City University of New York (CUNY). I am seeking participants for a research project entitled "Assessment Practices of School Psychologists." I am the Principal Investigator of this project, which is my dissertation. This is a research study examining school psychologists' testing practices.

Participation in this study involves completing a questionnaire, which takes about 20 minutes; the link to the online survey is provided at the end of this message. There are no risks to you in taking part; you do not have to provide your name or any other identifying information. Taking part is voluntary. You can choose not to complete the questionnaire.

If you would like a copy of the results of this study, please email me your address, and I will send you a copy in the future. In addition, if you choose to complete the study, you may enter your email address at the end of the questionnaire to be entered into a drawing to receive 1 of 3 $25 Amazon.com gift certificates. Email addresses will not be tied to participants' responses after their completion of the questionnaire.

If you have any questions about this study, you can contact me at swhitney@gc.cuny.edu, or my advisor, Georgiana Shick Tryon, at gtryon@gc.cuny.edu. If you have questions about your rights as a participant in this study, you can contact Kay Powell, IRB Administrator, The Graduate Center/City University of New York, (212) 817-7525, kpowell@gc.cuny.edu.

Thank you for your participation in the study.

http://www.surveymonkey.com/s/srwdissertation

Appendix B
Research Questionnaire

My name is Sarah Whitney, and I am a student in the School Psychology Ph.D. Program in the Educational Psychology Department at The Graduate Center of the City University of New York (CUNY). I am the Principal Investigator of this project, my dissertation, entitled "Assessment Practices of School Psychologists." This is a research study examining school psychologists' testing practices.

Participation in this study involves completing a questionnaire, which takes about 20 minutes. There are no risks to you in taking part; you do not have to provide your name or any other identifying information. Taking part is voluntary. You can choose not to complete the questionnaire.

If you would like a copy of the results of this study, please email me your address, and I will send you a copy in the future. In addition, if you choose to complete the study, you may enter your email address at the end of the questionnaire to be entered into a drawing to receive 1 of 3 $25 Amazon.com gift certificates. Email addresses will not be tied to participants' responses after their completion of the questionnaire.

If you have any questions about this study, you can contact me at swhitney@gc.cuny.edu, or my advisor, Georgiana Shick Tryon, at gtryon@gc.cuny.edu. If you have questions about your rights as a participant in this study, you can contact Kay Powell, IRB Administrator, The Graduate Center/City University of New York, (212) 817-7525, kpowell@gc.cuny.edu.

Thank you for your participation in the study.

Please complete the following background information questions.

Age

- ◯ 20-29
- ◯ 30-39
- ◯ 40-49
- ◯ 50-59
- ◯ 60-69
- ◯ 70+

Gender

- ◯ Male
- ◯ Female

Ethnicity

- ◯ Caucasian
- ◯ Asian
- ◯ African-American
- ◯ Hispanic
- ◯ Other (please specify)

Please enter the city and state in which you live.

City/Town:

State:

Please complete the following questions about your education.

What is the highest degree you have earned?

- ◯ Bachelors
- ◯ Masters
- ◯ Ph.D.
- ◯ Psy.D.
- ◯ Ed.D.
- ◯ Other (please specify)

Do you hold any of the following credentials? Please check all that apply.

- ☐ State certification
- ☐ National certification
- ☐ State license
- ☐ Other (please specify)

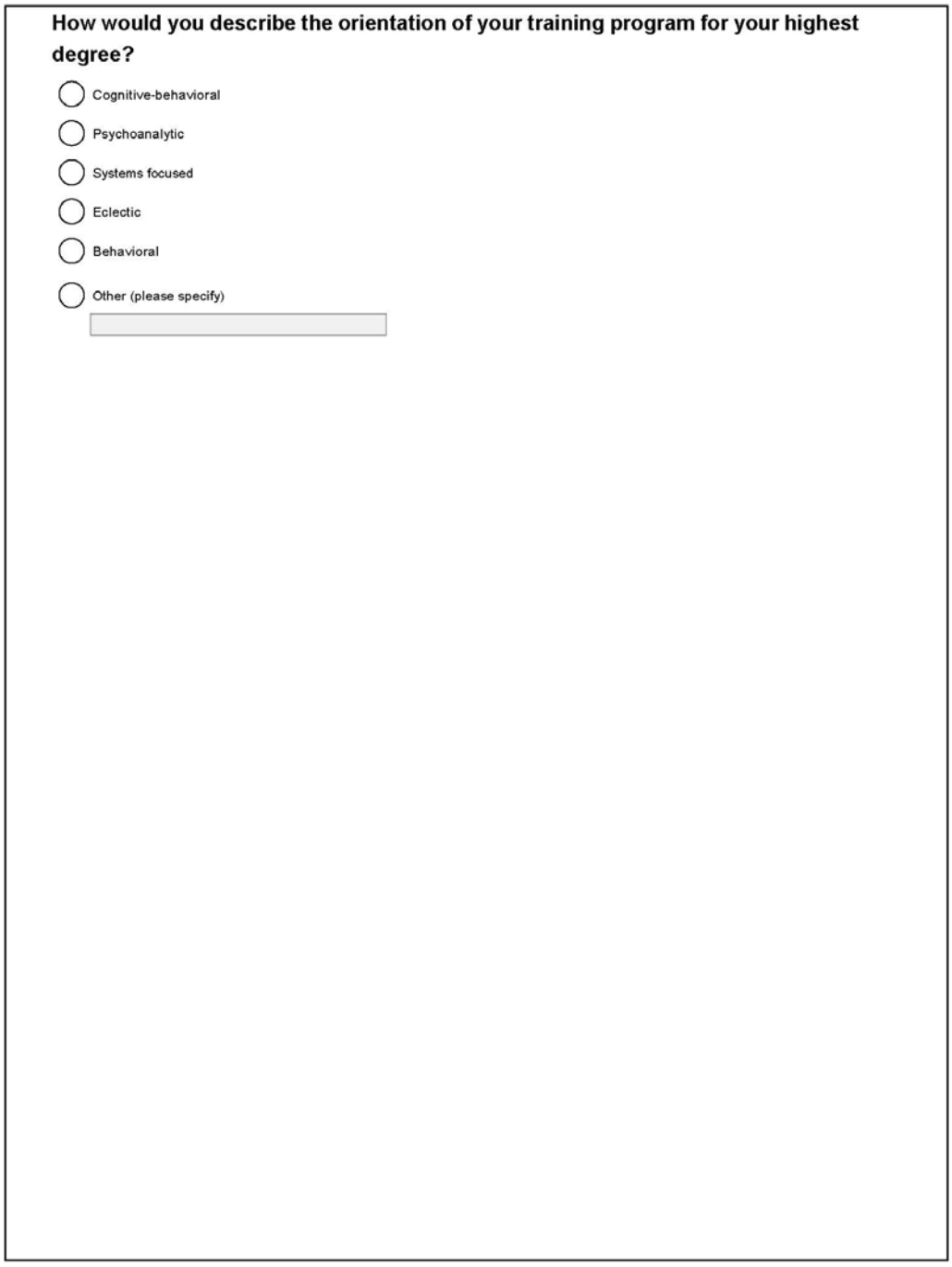
Where did you earn your highest degree?

University

State:

In what type of program did you earn your highest degree?

- ◯ School Psychology
- ◯ School-Clinical-Child Psychology
- ◯ Clinical Psychology
- ◯ Other (please specify)


How would you describe the orientation of your training program for your highest degree?
Cognitive-behavioral
Psychoanalytic
Systems focused
Eclectic
Behavioral
Other (please specify)

What is your primary workplace?

- ◯ Pre-school
- ◯ Elementary School
- ◯ Middle School
- ◯ High School
- ◯ Hospital
- ◯ Private Practice
- ◯ Other (please specify)

In what state is your primary workplace?

State:

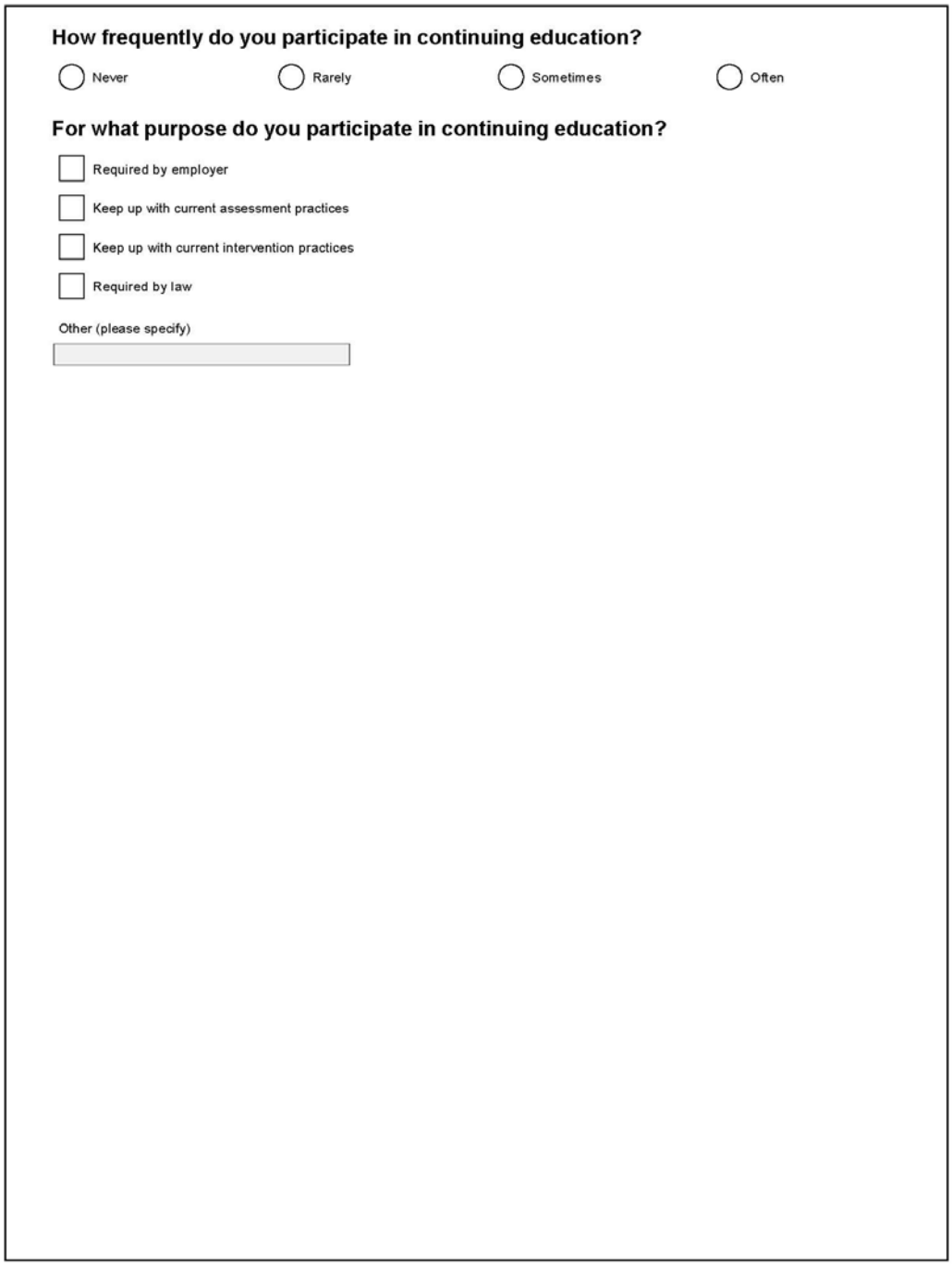
How many years have you been working as a school psychologist?

Please rank how much time you spent performing the following activities during the 2009-2010 school year, with 1 representing the most amount of time and 4 representing the least amount of time.

	1	2	3	4
Assessment/Testing	◯	◯	◯	◯
Consultation	◯	◯	◯	◯
Counseling	◯	◯	◯	◯
Crisis Intervention	◯	◯	◯	◯

Do you participate in continuing education for school psychologists?

- ◯ No
- ◯ Yes



How frequently do you participate in continuing education?

◯ Never ◯ Rarely ◯ Sometimes ◯ Often

For what purpose do you participate in continuing education?

☐ Required by employer

☐ Keep up with current assessment practices

☐ Keep up with current intervention practices

☐ Required by law

Other (please specify)

_____

Are you a member of APA?

- ◯ Yes
- ◯ No

Are you a member of NASP?

- ◯ Yes
- ◯ No

Please list any other professional organizations to which you belong.

For these questions, please think about your professional identity as a psychologist.

To what extent is your professional identity the result of your graduate school training and experiences?

- ◯ Not at all
- ◯ Very little
- ◯ Somewhat
- ◯ A lot

To what extent is your professional identity the result of your experiences since you left graduate school?

- ◯ Not at all
- ◯ Very little
- ◯ Somewhat
- ◯ A lot

For the following questions, please think about how much time you spent on campus during graduate school.

On average, how many hours a day did you spend on campus during graduate school (excluding internship)? ☐

On average, how many days a week did you spend on campus during graduate school (excluding internship)? ☐

As you know, when you choose a test to administer, you consider its psychometric properties, how convenient the test is to administer and interpret, and how much clinical judgment the test requires you to exercise.

<u>Now imagine you have to choose between two tests (you will be asked to make 13 of these choices in total):</u>

Test A has reliability coefficients of .90, correlates highly with measures of similar constructs, and has norms based on a sample very similar to the student you are testing. You have 100 copies of Test A on hand in your office, and it takes 10 minutes to administer and score. Test A items are ambiguous and allow students to respond freely. The manual provides suggestions but permits you to exercise your judgment in administration and scoring.

Test B has reliability coefficients of .90, correlates highly with measures of similar constructs, and has norms based on a sample very similar to the student you are testing. You have 100 copies of Test B on hand in your office, and it takes 10 minutes to administer and score. Test B items are standard and allow students to respond freely. The manual provides a standard administration protocol and suggestions for scoring that require you to exercise judgment in assigning scores.

	Test A	Test B
I would choose:	◯	◯

Now imagine you have to choose between the following two tests:

Test A has reliability coefficients of .90, correlates highly with measures of similar constructs, and has norms based on a sample very similar to the student you are testing. You have 100 copies of Test A on hand in your office, and it takes 10 minutes to administer and score. Test A items are standard and allow students to respond freely. The manual provides a standard administration protocol and suggestions for scoring that require you to exercise judgment in assigning scores.

Test B has reliability coefficients of .90, correlates highly with measures of similar constructs, and has norms based on a sample very similar to the student you are testing. You have 100 copies of Test B on hand in your office, and it takes 10 minutes to administer and score. Test B items are standard and students respond by marking one of several alternatives. The manual provides exact instructions for standard administration and scoring.

	Test A	Test B
I would choose:	◯	◯

Now imagine you have to choose between the following two tests:

Test A has reliability coefficients of .90, correlates highly with measures of similar constructs, and has norms based on a sample very similar to the student you are testing. You have 100 copies of Test A on hand in your office, and it takes 10 minutes to administer and score. Test A items are standard and allow students to respond freely. The manual provides a standard administration protocol and suggestions for scoring that require you to exercise judgment in assigning scores.

Test B has reliability coefficients of .80, correlates moderately with measures of similar constructs, and has norms that reflect the general school population. You have 100 copies of Test B on hand in your office, and it takes 10 minutes to administer and score. Test B items are standard and allow students to respond freely. The manual provides a standard administration protocol and suggestions for scoring that require you to exercise judgment in assigning scores.

	Test A	Test B
I would choose:	◯	◯

Now imagine you have to choose between the following two tests:

Test A has reliability coefficients of .90, correlates highly with measures of similar constructs, and has norms based on a sample very similar to the student you are testing. You have 100 copies of Test A on hand in your office, and it takes 10 minutes to administer and score. Test A items are standard and allow students to respond freely. The manual provides a standard administration protocol and suggestions for scoring that require you to exercise judgment in assigning scores.

Test B has reliability coefficients of .80, correlates moderately with measures of similar constructs, and has norms that reflect the general school population. You purchase Test B protocols for $25 per a pack of 5, and it takes 20 minutes to administer and score. Test B items are standard and students respond by marking one of several alternatives. The manual provides exact instructions for standard administration and scoring.

	Test A	Test B
I would choose:	◯	◯

Now imagine you have to choose between the following two tests:

Test A has reliability coefficients of .90, correlates highly with measures of similar constructs, and has norms based on a sample very similar to the student you are testing. You have 100 copies of Test A on hand in your office, and it takes 10 minutes to administer and score. Test A items are standard and students respond by marking one of several alternatives. The manual provides exact instructions for standard administration and scoring.

Test B has reliability coefficients of .80, correlates moderately with measures of similar constructs, and has norms that reflect the general school population. You have 100 copies of Test B on hand in your office, and it takes 10 minutes to administer and score. Test B items are ambiguous and allow students to respond freely. The manual provides suggestions but permits you to exercise your judgment in administration and scoring.

	Test A	Test B
I would choose:	◯	◯

Now imagine you have to choose between the following two tests:

Test A has reliability coefficients of .90, correlates highly with measures of similar constructs, and has norms based on a sample very similar to the student you are testing. You purchase Test A protocols for $25 per a pack of 5, and it takes 20 minutes to administer and score. Test A items are ambiguous and allow students to respond freely. The manual provides suggestions but permits you to exercise your judgment in administration and scoring.

Test B has reliability coefficients of .80, correlates moderately with measures of similar constructs, and has norms that reflect the general school population. You have 100 copies of Test B on hand in your office, and it takes 10 minutes to administer and score. Test B items are ambiguous and allow students to respond freely. The manual provides suggestions but permits you to exercise your judgment in administration and scoring.

	Test A	Test B
I would choose:	◯	◯

Now imagine you have to choose between the following two tests:

Test A has reliability coefficients of .90, correlates highly with measures of similar constructs, and has norms based on a sample very similar to the student you are testing. You purchase Test A protocols for $25 per a pack of 5, and it takes 20 minutes to administer and score. Test A items are standard and allow students to respond freely. The manual provides a standard administration protocol and suggestions for scoring that require you to exercise judgment in assigning scores.

Test B has reliability coefficients of .90, correlates highly with measures of similar constructs, and has norms based on a sample very similar to the student you are testing. You purchase Test B protocols for $25 per a pack of 5, and it takes 20 minutes to administer and score. Test B items are standard and students respond by marking one of several alternatives. The manual provides exact instructions for standard administration and scoring.

	Test A	Test B
I would choose:	◯	◯

Now imagine you have to choose between the following two tests:

Test A has reliability coefficients of .90, correlates highly with measures of similar constructs, and has norms based on a sample very similar to the student you are testing. You purchase Test A protocols for $25 per a pack of 5, and it takes 20 minutes to administer and score. Test A items are standard and students respond by marking one of several alternatives. The manual provides exact instructions for standard administration and scoring.

Test B has reliability coefficients of .80, correlates moderately with measures of similar constructs, and has norms that reflect the general school population. You have 100 copies of Test B on hand in your office, and it takes 10 minutes to administer and score. Test B items are standard and allow students to respond freely. The manual provides a standard administration protocol and suggestions for scoring that require you to exercise judgment in assigning scores.

	Test A	Test B
I would choose:	◯	◯

Now imagine you have to choose between the following two tests:

Test A has reliability coefficients of .80, correlates moderately with measures of similar constructs, and has norms that reflect the general school population. You have 100 copies of Test A on hand in your office, and it takes 10 minutes to administer and score. Test A items are ambiguous and allow students to respond freely. The manual provides suggestions but permits you to exercise your judgment in administration and scoring.

Test B has reliability coefficients of .80, correlates moderately with measures of similar constructs, and has norms that reflect the general school population. You have 100 copies of Test B on hand in your office, and it takes 10 minutes to administer and score. Test B items are standard and allow students to respond freely. The manual provides a standard administration protocol and suggestions for scoring that require you to exercise judgment in assigning scores.

	Test A	Test B
I would choose:	◯	◯

Now imagine you have to choose between the following two tests:

Test A has reliability coefficients of .80, correlates moderately with measures of similar constructs, and has norms that reflect the general school population. You have 100 copies of Test A on hand in your office, and it takes 10 minutes to administer and score. Test A items are standard and allow students to respond freely. The manual provides a standard administration protocol and suggestions for scoring that require you to exercise judgment in assigning scores.

Test B has reliability coefficients of .80, correlates moderately with measures of similar constructs, and has norms that reflect the general school population. You purchase Test B protocols for $25 per a pack of 5, and it takes 20 minutes to administer and score. Test B items are ambiguous and allow students to respond freely. The manual provides suggestions but permits you to exercise your judgment in administration and scoring.

	Test A	Test B
I would choose:	◯	◯

Now imagine you have to choose between the following two tests:

Test A has reliability coefficients of .80, correlates moderately with measures of similar constructs, and has norms that reflect the general school population. You purchase Test A protocols for $25 per a pack of 5, and it takes 20 minutes to administer and score. Test A items are ambiguous and allow students to respond freely. The manual provides suggestions but permits you to exercise your judgment in administration and scoring.

Test B has reliability coefficients of .80, correlates moderately with measures of similar constructs, and has norms that reflect the general school population. You purchase Test B protocols for $25 per a pack of 5, and it takes 20 minutes to administer and score. Test B items are standard and allow students to respond freely. The manual provides a standard administration protocol and suggestions for scoring that require you to exercise judgment in assigning scores.

	Test A	Test B
I would choose:	◯	◯

Now imagine you have to choose between the following two tests:

Test A has reliability coefficients of .90, correlates highly with measures of similar constructs, and has norms based on a sample very similar to the student you are testing. You have 100 copies of Test A on hand in your office, and it takes 10 minutes to administer and score. Test A items are standard and allow students to respond freely. The manual provides a standard administration protocol and suggestions for scoring that require you to exercise judgment in assigning scores.

Test B has reliability coefficients of .80, correlates moderately with measures of similar constructs, and has norms that reflect the general school population. You have 100 copies of Test B on hand in your office, and it takes 10 minutes to administer and score. Test B items are ambiguous and allow students to respond freely. The manual provides suggestions but permits you to exercise your judgment in administration and scoring.

	Test A	Test B
I would choose:	◯	◯

Now imagine you have to choose between the following two tests:

Test A has reliability coefficients of .90, correlates highly with measures of similar constructs, and has norms based on a sample very similar to the student you are testing. You have 100 copies of Test A on hand in your office, and it takes 10 minutes to administer and score. Test A items are standard and students respond by marking one of several alternatives. The manual provides exact instructions for standard administration and scoring.

Test B has reliability coefficients of .80, correlates moderately with measures of similar constructs, and has norms that reflect the general school population. You have 100 copies of Test B on hand in your office, and it takes 10 minutes to administer and score. Test B items are standard and allow students to respond freely. The manual provides a standard administration protocol and suggestions for scoring that require you to exercise judgment in assigning scores.

	Test A	Test B
I would choose:	◯	◯

Please mark which of the test attributes is more important to you. The closer your mark is to the attribute that is in bold type on top of the choices, the more important that attribute is to you.

Psychometric Properties				**Convenience of Administration**
Most Important	More Important	About Equal	More Important	Most Important
◯	◯	◯	◯	◯

Please mark which of the test attributes is more important to you. The closer your mark is to the attribute that is in bold type on top of the choices, the more important that attribute is to you.

Psychometric Properties				**Clinical Judgment**
Most Important	More Important	About Equal	More Important	Most Important
◯	◯	◯	◯	◯

Please mark which of the test attributes is more important to you. The closer your mark is to the attribute that is in bold type on top of the choices, the more important that attribute is to you.

Convenience of Administration				**Clinical Judgment**
Most Important	More Important	About Equal	More Important	Most Important
◯	◯	◯	◯	◯

Have you heard of the following assessment measures?

	Yes	No
Achenbach Child Behavior Checklist	◯	◯
BASC-2	◯	◯
Bender-Gestalt	◯	◯
Children's Apperception Test	◯	◯
Draw a Person	◯	◯
House Tree Person	◯	◯
Kinetic Family Drawing	◯	◯
Robert's Apperception Test	◯	◯
Rorschach	◯	◯
Sentence Completion	◯	◯
Thematic Apperception Test	◯	◯

Have you used the following assessment measures as part of an assessment you conducted in the 2009-2010 school year?

	Yes	No
Achenbach Child Behavior Checklist	◯	◯
BASC-2	◯	◯
Bender-Gestalt	◯	◯
Children's Apperception Test	◯	◯
Draw a Person	◯	◯
House Tree Person	◯	◯
Kinetic Family Drawing	◯	◯
Robert's Apperception Test	◯	◯
Rorschach	◯	◯
Sentence Completion	◯	◯
Thematic Apperception Test	◯	◯

Of the following tests, please choose and rank the three you feel most confident administering and interpreting the results of. Please assign a 1 to the test you are most confident in your ability to administer and interpret, 2 to the next test you are most confident in your ability to administer and interpret, and 3 to the last test you are most confident in your ability to administer and interpret.

	1	2	3
Achenbach Child Behavior Checklist	◯	◯	◯
BASC-2	◯	◯	◯
Bender-Gestalt	◯	◯	◯
Children's Apperception Test	◯	◯	◯
Draw a Person	◯	◯	◯
House Tree Person	◯	◯	◯
Kinetic Family Drawing	◯	◯	◯
Robert's Apperception Test	◯	◯	◯
Rorschach	◯	◯	◯
Sentence Completion	◯	◯	◯
Thematic Apperception Test	◯	◯	◯

Of the following tests, please choose and rank the three you consider to have the best psychometric properties. Please assign a 1 to the test you think has the best psychometric properties, a 2 to the test you think has the next best psychometric properties, and a 3 to the last test you think has the best psychometric properties.

	1	2	3
Achenbach Child Behavior Checklist	◯	◯	◯
BASC-2	◯	◯	◯
Bender-Gestalt	◯	◯	◯
Children's Apperception Test	◯	◯	◯
Draw a Person	◯	◯	◯
House Tree Person	◯	◯	◯
Kinetic Family Drawing	◯	◯	◯
Robert's Apperception Test	◯	◯	◯
Rorschach	◯	◯	◯
Sentence Completion	◯	◯	◯
Thematic Apperception Test	◯	◯	◯

Of the following tests, please choose and rank the three you consider to be most useful. Please assign a 1 to the test you think is the most useful, a 2 to the test you think is the next most useful, and a 3 to the last test you think is most useful.

	1	2	3
Achenbach Child Behavior Checklist	◯	◯	◯
BASC-2	◯	◯	◯
Bender-Gestalt	◯	◯	◯
Children's Apperception Test	◯	◯	◯
Draw a Person	◯	◯	◯
House Tree Person	◯	◯	◯
Kinetic Family Drawing	◯	◯	◯
Robert's Apperception Test	◯	◯	◯
Rorschach	◯	◯	◯
Sentence Completion	◯	◯	◯
Thematic Apperception Test	◯	◯	◯

Of the following test characteristics, please choose and rank the three you consider to be most important. Please assign a 1 to the characteristic you think is the most important, a 2 to the characteristic you think is the next most important, and a 3 to the last characteristic you think is most important.

	1	2	3
Directly addresses the referral question.	◯	◯	◯
Has good validity.	◯	◯	◯
Is easy to score.	◯	◯	◯
Yields information not available from other measures.	◯	◯	◯
Has strong psychometric properties.	◯	◯	◯
Has an appropriate norming sample.	◯	◯	◯
Has good test-retest reliability.	◯	◯	◯
Yields results that are easy to interpret.	◯	◯	◯
Is a required assessment measure in your workplace.	◯	◯	◯
Is part of your standard assessment battery.	◯	◯	◯
Requires judgment unique to your training.	◯	◯	◯

For this question, please imagine that you are required to administer and interpret <u>a test that does not have norms that are representative of the population you are using the test for.</u> Please mark the extent to which each of the words below applies to how you are feeling currently:

	Does not apply at all	Does not apply	Does not apply somewhat	Undecided	Applies somewhat	Applies	Applies very much
Uneasy	◯	◯	◯	◯	◯	◯	◯
Uncomfortable	◯	◯	◯	◯	◯	◯	◯
Tense	◯	◯	◯	◯	◯	◯	◯
Bothered	◯	◯	◯	◯	◯	◯	◯
Concerned	◯	◯	◯	◯	◯	◯	◯

For this question, please imagine that you are required to administer and interpret <u>a test that you judge to have less than ideal psychometric evidence to support it, i.e., less than ideal reliability and validity.</u> Please mark the extent to which each of the words below applies to how you are feeling currently:

	Does not apply at all	Does not apply	Does not apply somewhat	Undecided	Applies somewhat	Applies	Applies very much
Uneasy	◯	◯	◯	◯	◯	◯	◯
Uncomfortable	◯	◯	◯	◯	◯	◯	◯
Tense	◯	◯	◯	◯	◯	◯	◯
Bothered	◯	◯	◯	◯	◯	◯	◯
Concerned	◯	◯	◯	◯	◯	◯	◯

For this question, please imagine that you are required to administer and interpret a test that does not have standardized administration or scoring procedures. Please mark the extent to which each of the words below applies to how you are feeling currently:

	Does not apply at all	Does not apply	Does not apply somewhat	Undecided	Applies somewhat	Applies	Applies very much
Uneasy	◯	◯	◯	◯	◯	◯	◯
Uncomfortable	◯	◯	◯	◯	◯	◯	◯
Tense	◯	◯	◯	◯	◯	◯	◯
Bothered	◯	◯	◯	◯	◯	◯	◯
Concerned	◯	◯	◯	◯	◯	◯	◯

For this question, please imagine that you are required to administer and interpret a test that requires you to use considerable clinical judgment. Please mark the extent to which each of the words below applies to how you are feeling currently:

	Does not apply at all	Does not apply	Does not apply somewhat	Undecided	Applies somewhat	Applies	Applies very much
Uneasy	◯	◯	◯	◯	◯	◯	◯
Uncomfortable	◯	◯	◯	◯	◯	◯	◯
Tense	◯	◯	◯	◯	◯	◯	◯
Bothered	◯	◯	◯	◯	◯	◯	◯
Concerned	◯	◯	◯	◯	◯	◯	◯

Here are a number of characteristics that may or may not apply to you. Please indicate the extent to which you agree or disagree with each statement.

I see myself as someone who:

	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
Is talkative	◯	◯	◯	◯	◯
Tends to find fault with others	◯	◯	◯	◯	◯
Does a thorough job	◯	◯	◯	◯	◯
Is depressed, blue	◯	◯	◯	◯	◯
Is original, comes up with new ideas	◯	◯	◯	◯	◯
Is reserved	◯	◯	◯	◯	◯
Is helpful and unselfish with others	◯	◯	◯	◯	◯
Can be somewhat careless	◯	◯	◯	◯	◯
Is relaxed, handles stress well	◯	◯	◯	◯	◯
Is curious about many different things	◯	◯	◯	◯	◯
Is full of energy	◯	◯	◯	◯	◯
Starts quarrels with others	◯	◯	◯	◯	◯
Is a reliable worker	◯	◯	◯	◯	◯
Can be tense	◯	◯	◯	◯	◯

I see myself as someone who:

	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
Is ingenious, a deep thinker	◯	◯	◯	◯	◯
Generates a lot of enthusiasm	◯	◯	◯	◯	◯
Has a forgiving nature	◯	◯	◯	◯	◯
Tends to be disorganized	◯	◯	◯	◯	◯
Worries a lot	◯	◯	◯	◯	◯
Has an active imagination	◯	◯	◯	◯	◯
Tends to be quiet	◯	◯	◯	◯	◯
Is generally trusting	◯	◯	◯	◯	◯
Tends to be lazy	◯	◯	◯	◯	◯
Is emotionally stable, not easily upset	◯	◯	◯	◯	◯
Is inventive	◯	◯	◯	◯	◯
Has an assertive personality	◯	◯	◯	◯	◯
Can be cold and aloof	◯	◯	◯	◯	◯
Perseveres until the task is finished	◯	◯	◯	◯	◯

I see myself as someone who:

	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
Can be moody	◯	◯	◯	◯	◯
Values artistic, aesthetic experiences	◯	◯	◯	◯	◯
Is sometimes shy, inhibited	◯	◯	◯	◯	◯
Is considerate and kind to almost everyone	◯	◯	◯	◯	◯
Does things efficiently	◯	◯	◯	◯	◯
Remains calm in tense situations	◯	◯	◯	◯	◯
Prefers work that is routine	◯	◯	◯	◯	◯
Is outgoing, sociable	◯	◯	◯	◯	◯
Is sometimes rude to others	◯	◯	◯	◯	◯
Makes plans and follows through with them	◯	◯	◯	◯	◯
Gets nervous easily	◯	◯	◯	◯	◯
Likes to reflect, play with ideas	◯	◯	◯	◯	◯
Has few artistic interests	◯	◯	◯	◯	◯
Likes to cooperate with others	◯	◯	◯	◯	◯

I see myself as someone who:

	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
Is easily distracted	◯	◯	◯	◯	◯
Is sophisticated in art, music, or literature	◯	◯	◯	◯	◯

Here are a number of characteristics that may or may not apply to you. Please indicate the extent to which you agree or disagree with each statement.

Please mark to what extent you agree with each statement.

	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
I will be able to achieve most of the goals that I have set for myself.	◯	◯	◯	◯	◯
When facing difficult tasks, I am certain that I will accomplish them.	◯	◯	◯	◯	◯
In general, I think that I can obtain outcomes that are important to me.	◯	◯	◯	◯	◯
I believe I can succeed at most any endeavor to which I set my mind.	◯	◯	◯	◯	◯
I will be able to successfully overcome many challenges.	◯	◯	◯	◯	◯
I am confident that I can perform effectively on many different tasks.	◯	◯	◯	◯	◯
Compared to other people, I can do most tasks very well.	◯	◯	◯	◯	◯
Even when things are tough, I can perform quite well.	◯	◯	◯	◯	◯

How well can you:

	Not well at all	Not too well	Neutral	Pretty well	Very well
Evaluate the psychometric properties of tests?	◯	◯	◯	◯	◯
Change or add tests of other assessment procedures as a result of the information you obtain early in the assessment process?	◯	◯	◯	◯	◯
Give clinician administered assessment instruments?	◯	◯	◯	◯	◯
Follow the legal and ethical standards of school psychology in practice?	◯	◯	◯	◯	◯
Follow standardized procedures when using assessment tools?	◯	◯	◯	◯	◯
Examine school records?	◯	◯	◯	◯	◯
Administer visual-motor tests?	◯	◯	◯	◯	◯
Administer social, emotional, and behavior measures?	◯	◯	◯	◯	◯
Score assessment measures?	◯	◯	◯	◯	◯

How well can you:

	Not well at all	Not too well	Neutral	Pretty well	Very well
Administer intelligence tests?	◯	◯	◯	◯	◯
Evaluate the appropriateness of the norm group when interpreting the test scores of an individual or group?	◯	◯	◯	◯	◯
Follow the steps in the assessment process?	◯	◯	◯	◯	◯
Write to effectively communicate the most important points of a psychoeducational assessment?	◯	◯	◯	◯	◯
Understand the consequences of assessment-related decisions?	◯	◯	◯	◯	◯
Administer achievement tests?	◯	◯	◯	◯	◯
Interpret comprehensive assessment results for decision making purposes?	◯	◯	◯	◯	◯
Administer adaptive behavior measures?	◯	◯	◯	◯	◯
Choose assessment instruments for addressing the referral concern(s)?	◯	◯	◯	◯	◯

Additional Comments and/or Suggestions:

Thank you for completing the survey! Please click on the link below into your web browser to enter your email address into a drawing to win one of three $25 Amazon.com gift certificates. Note: this link will take you to a separate website, and your email address will not be tied to your responses to this questionnaire. Click here to enter your email address

Appendix C

Characteristics of Hypothetical Tests

Psychometric properties

Excellent	The test has reliability coefficients of .90, correlates highly with measures of similar constructs, and has norms based on a sample very similar to the student you are testing.
Adequate	The test has reliability coefficients of .80, correlates moderately with measures of similar constructs, and has norms that reflect the general school population.

Convenience

Very	You have 100 copies on hand in your office, and it takes 10 minutes to administer and score.
Moderate	You purchase test protocols for $25 per a pack of 5, and it takes 20 minutes to administer and score.

Clinical Judgment

A lot	Items are ambiguous and allow students to respond freely. The manual provides suggestions but permits you to exercise your judgment in administration and scoring.
Moderate	Items are standard and allow students to respond freely. The manual provides a standard administration protocol and suggestions for scoring that require you to exercise judgment in assigning scores.
Very little	Items are standard and students respond by marking one of several alternatives. The manual provides exact instructions for standard administration and scoring.

Appendix D
List of Hypothetical Tests

Hypothetical Test #	Psychometric Properties	Convenience	Clinical Judgment
1	Excellent	Very	A lot
2	Excellent	Very	Moderate
3	Excellent	Very	Very little
4	Excellent	Moderately	A lot
5	Excellent	Moderately	Moderate
6	Excellent	Moderately	Very little
7	Adequate	Very	A lot
8	Adequate	Very	Moderate
9	Adequate	Very	Very little
10	Adequate	Moderately	A lot
11	Adequate	Moderately	Moderate
12	Adequate	Moderately	Very little

Appendix E

Breakdown of Hypothetical Tests by Questionnaire

Pair #	Test #1	Test #2	Questionnaire #		Pair #	Test #1	Test #2	Questionnaire #
1	1	2	1		1	1	5	4
2	2	3	1		2	1	11	4
3	2	8	1		3	2	4	4
4	2	12	1		4	3	4	4
5	3	7	1		5	3	10	4
6	4	7	1		6	4	9	4
7	5	6	1		7	5	8	4
8	6	8	1		8	5	10	4
9	7	8	1		9	7	9	4
10	8	10	1		10	7	12	4
11	10	11	1		11	9	12	4
12	2	7	1		12	2	7	4
13	3	8	1		13	3	8	4
1	1	3	2		1	1	6	5
2	1	10	2		2	1	9	5
3	2	5	2		3	2	6	5
4	2	11	2		4	3	6	5
5	3	11	2		5	3	12	5
6	4	8	2		6	4	5	5
7	4	11	2		7	5	7	5
8	5	12	2		8	5	11	5
9	6	11	2		9	6	12	5
10	8	12	2		10	8	9	5
11	10	12	2		11	9	10	5
12	2	7	2		12	2	7	5
13	3	8	2		13	3	8	5
1	1	4	3		1	1	7	6
2	1	8	3		2	1	12	6
3	2	9	3		3	2	7	6
4	3	5	3		4	2	10	6
5	3	9	3		5	3	8	6
6	4	10	3		6	4	6	6
7	5	9	3		7	4	12	6
8	6	7	3		8	6	9	6
9	6	10	3		9	7	10	6
10	7	11	3		10	8	11	6
11	11	12	3		11	9	11	6
12	2	7	3					
13	3	8	3					

Pair #	Test #1	Test #2	Questionnaire #		Pair #	Test #1	Test #2	Questionnaire #
1	1	2	7		1	2	3	10
2	2	8	7		2	2	12	10
3	3	7	7		3	4	7	10
4	5	6	7		4	6	8	10
5	7	8	7		5	8	10	10
6	10	11	7		6	2	7	10
7	3	8	7		7	1	7	10
8	1	11	7		8	3	8	10
9	3	4	7		9	4	12	10
10	4	9	7		10	7	10	10
11	5	10	7		11	9	11	10
12	7	12	7					
13	2	7	7					
1	1	3	8		1	1	10	11
2	2	5	8		2	2	11	11
3	3	11	8		3	4	8	11
4	4	11	8		4	5	12	11
5	6	11	8		5	8	12	11
6	10	12	8		6	2	7	11
7	3	8	8		7	1	6	11
8	1	9	8		8	2	6	11
9	3	6	8		9	3	12	11
10	4	5	8		10	5	7	11
11	5	11	8		11	6	12	11
12	8	9	8		12	9	10	11
13	2	7	8		13	3	8	11
1	1	4	9		1	1	8	12
2	2	9	9		2	3	5	12
3	3	9	9		3	4	10	12
4	5	9	9		4	6	7	12
5	6	10	9		5	7	11	12
6	11	12	9		6	2	7	12
7	3	8	9		7	1	5	12
8	1	12	9		8	2	4	12
9	2	7	9		9	3	10	12
10	2	10	9		10	5	8	12
11	4	6	9		11	7	9	12
12	6	9	9		12	9	12	12
					13	3	8	12

Appendix F
Results of t-Test Analysis

Results of t-Test Comparing Usage of the Achenbach

Comparison Variable	Difference Between Groups	t Value
Days a Week Spent on Campus ($N = 107$)	-0.23	-1.09
Hours a Day Spent on Campus ($N = 103$)	-1.03	-1.41
Professional Identity as Result of Graduate School ($N = 111$)	-0.02	-0.20
Professional Identity as Result of Post-Graduate School ($N = 111$)	0.23	1.70
Total Cognitive Dissonance ($N = 111$)	-5.04	-1.22
General Self-Efficacy ($N = 111$)	1.32	1.90
SP Self-Efficacy ($N = 111$)	2.34	1.84
Extraversion ($N = 111$)	-2.01	-1.51

Results of t-Test Comparing Usage of the Bender-Gestalt

Comparison Variable	Difference Between Groups	t Value
Days a Week Spent on Campus ($N = 105$)	0.27	1.19
Hours a Day Spent on Campus ($N = 102$)	0.14	0.19
Professional Identity as Result of Graduate School ($N = 109$)	0.08	0.67
Professional Identity as Result of Post-Graduate School ($N = 109$)	-0.07	-0.48
Total Cognitive Dissonance ($N = 109$)	7.96	1.79
General Self-Efficacy ($N = 109$)	-0.58	-0.84
SP Self-Efficacy ($N = 109$)	-1.48	-1.24
Extraversion ($N = 109$)	-1.68	-1.23

Results of t-Test Comparing Usage of the Draw a Person

Comparison Variable	Difference Between Groups	t Value
Days a Week Spent on Campus ($N = 105$)	0.11	0.49
Hours a Day Spent on Campus ($N = 102$)	0.14	0.18
Professional Identity as Result of Graduate School ($N = 109$)	0.21	1.75
Professional Identity as Result of Post-Graduate School ($N = 109$)	0.04	0.25
Total Cognitive Dissonance ($N = 109$)	9.49	2.16*
General Self-Efficacy ($N = 109$)	-0.64	-0.89
SP Self-Efficacy ($N = 109$)	-2.17	-1.72
Extraversion ($N = 109$)	-0.63	-0.44

* $p < .05$.

Results of t-Test Comparing Usage of the House Tree Person

Comparison Variable	Difference Between Groups	t Value
Days a Week Spent on Campus ($N = 103$)	0.58	2.51*
Hours a Day Spent on Campus ($N = 100$)	1.24	1.56
Professional Identity as Result of Graduate School ($N = 107$)	0.32	2.76**
Professional Identity as Result of Post-Graduate School ($N = 107$)	-0.26	-1.89
Total Cognitive Dissonance ($N = 107$)	3.30	0.73
General Self-Efficacy ($N = 107$)	-1.15	-1.58
SP Self-Efficacy ($N = 107$)	-2.71	-2.12*
Extraversion ($N = 107$)	-1.63	-1.13

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

Results of t-Test Comparing Usage of the Kinetic Family Drawing

Comparison Variable	Difference Between Groups	t Value
Days a Week Spent on Campus ($N = 103$)	0.15	0.63
Hours a Day Spent on Campus ($N = 100$)	-0.06	-0.08
Professional Identity as Result of Graduate School ($N = 107$)	0.17	1.41
Professional Identity as Result of Post-Graduate School ($N = 107$)	-0.10	-0.64
Total Cognitive Dissonance ($N = 107$)	4.91	1.07
General Self-Efficacy ($N = 107$)	-0.99	-1.31
SP Self-Efficacy ($N = 107$)	-1.49	-1.14
Extraversion ($N = 107$)	-0.69	-0.47

Results of t-Test Comparing Usage of Sentence Completion

Comparison Variable	Difference Between Groups	t Value
Days a Week Spent on Campus ($N = 107$)	0.30	1.36
Hours a Day Spent on Campus ($N = 104$)	0.86	1.21
Professional Identity as Result of Graduate School ($N = 111$)	0.12	1.10
Professional Identity as Result of Post-Graduate School ($N = 111$)	-0.20	-1.52
Total Cognitive Dissonance ($N = 111$)	7.89	1.85
General Self-Efficacy ($N = 111$)	-0.80	-1.23
SP Self-Efficacy ($N = 111$)	-1.49	-1.30
Extraversion ($N = 111$)	-0.84	-0.64

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *The standards for educational and psychological testing* (2nd ed.). Washington, DC: American Educational Research Association.

American Psychological Association. (n.d.). Archival description of the specialty school psychology. Retrieved from http://www.indiana.edu/~div16/goals.html

American Psychological Association (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57,* 1060-1073. doi: 10.1037//0003-066X.57.12.1060

Anderson, C. J. (2009). Categorical data analysis with a psychometric twist. In R. E. Millsap & A. Maydeau-Olivares (Eds.), *The Sage Handbook of quantitative methods in psychology* (pp. 311-336). Thousand Oaks, CA: Sage.

Baker, T.B., McFall, R.M., & Shoham, V. (2009). Current status and future prospects of clinical psychology: Toward a scientifically principled approach to mental and behavioral health care. *Psychological Science in the Public Interest, 9,* 67-103. doi: 10.1111/j.1539-6053.2009.01036.x

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84,* 191-215. doi: 10.1037/0033-295X.84.2.191

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W.H. Freeman and Company.

Barker, M. (1938). The duties of a school psychologist. *Journal of Consulting Psychology, 2,* 180-183. doi: 10.1037/h0063459

Bell, I., & Mellor, D. (2009). Clinical judgements: Research and practice. *Australian Psychologist, 44,* 112-121. doi: 10.1080/00050060802550023

Bellak, L., & Bellak, S.S. (1949). *The Children's Apperception Test.* New York: C.P.S.

Bramlett, R.K., Murphy, J.J., Johnson, J., Wallingford, L., & Hall, J.D. (2002). Contemporary practices in school psychology: A national survey of roles and referral problems. *Psychology in the Schools, 39,* 327-335. doi: 10.1002/pits.10022

Bruss, K.V., & Kopala, M. (1993). Graduate school training in psychology: Its impact upon the development of professional identity. *Psychotherapy, 30,* 685-691.

Burke, S.M., Sparkes, A.C., & Allen-Collinson, J. (2008). High altitude climbers as ethnomethodologists making sense of cognitive dissonance: Ethnographic insights from an attempt to scale Mt. Everest. *The Sport Psychologist, 22,* 336-355.

Cabana, M.D., Rand, C.S., Powe, N.R., Wu, A.W., Wilson, M.H., Abboud, P.A., & Rubin, H.R. (1999). Why don't physicians follow clinical practice guidelines? A framework for improvement. *Journal of the American Medical Association, 282,* 1458-65. doi: 10.1111/j.1478-5153.2007.00227.x

Camara W.J., Nathan, J.S., and Puente, A.E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice, 31,* 141-154. doi: I0.1037//0735-7028.31.2.I41

Cashel, M.L. (2002). Child and adolescent psychological assessment: Current clinical practices and the impact of managed care. *Professional Psychology: Research and Practice, 33,* 446-453. doi: 10.1037//0735-7028.33.5.446

Chen, G., Gully, S.M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods, 4,* 62-83. doi: 10.1177/109442810141004

Chiou, W.B., & Wan, C.S. (2007). Using cognitive dissonance to induce adolescents' escaping from the claw of online gaming: The roles of personal responsibility and justification of cost. *CyberPsychology & Behavior, 10,* 663-670. doi: 10.1089/cpb.2007.9972

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159. doi: 10.1037/0033-2909.112.1.155

Colbeck, C.L. (2008). Professional identity development theory and doctoral education. *New Directions for Teaching and Learning, 113,* 9-16. doi: 10.1002/tl

Curtis, M.J., Hunley, S.A., & Grier, J.E.C. (2002). Relationships among the professional practices and demographic characteristics of school psychologists. *School Psychology Review, 31,* 30-42.

Curtis, M.J., Hunley, S.A., Walker, K.J., & Baker, A.C. (1999). Demographic characteristics and professional practices in school psychology. *School Psychology Review, 28,* 104-116.

Curtis, M.J., Lopez, A.D., Castillo, J.M., Batsche, G.M., Minch, D., & Smith, J.C. (2008). The status of school psychology: Demographic characteristics, employment conditions, professional practices, and continuing professional development. *Communiqué, 36(5),* 27-29.

Davis, J.M., & Sandoval, J. (1992). School psychologists' personalities: Award winners contrasted with a random sample. *Professional Psychology: Research and Practice, 23,* 418-420. doi: 10.1037/0735-7028.23.5.418

Digman, J.M. (1989). Five robust trait dimensions: Development, stability, and utility. *Journal of Personality, 57*, 195-214. DOI: 10.1111/1467-6494.ep8972652

Eddy, D. M. (2005). Evidence-based medicine: A unified approach. *Health Affairs, 24,* 9-17. doi: 10.B77/hlthaff.24.1.9

Elliott, A.J., & Devine, P.G. (1994). On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology, 67,* 382-394.

Eysenck, H. J. (1990). Biological dimensions of personality. In L. A. Pervin (Ed.), Handbook of personality: Theory and research (pp. 244–276). New York: Guilford Press.

Fagan, T.K., & Wise, P.S. (2007). *School psychology: past, present, and future* (3rd ed.). Bethesda, MD: National Association of School Psychologists.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, California: Stanford University Press.

Festinger, L., & Carlsmith, J.M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology, 58,* 203-210. doi: 10.1037/h0041593

Fiske, D.W. (1949). Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal Social Psychology, 44,* 329-344.

Fletcher, J.M., Francis, D.J., Morris, R.D., & Lyon, G.R. (2005). Evidence-based assessment of learning disabilities in children and adolescents. *Journal of Clinical*

Child and Adolescent Psychology, 34, 506-522. doi: 10.1207/s15374424jccp3403_7

Garb, H.N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105,* 387-396. doi: 10.1037/0033-2909.105.3.387

Garb, H.N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1,* 67-89. doi: 10.1146/annurev.clinpsy.1.102803.143810

Garb, H.N., Wood, J.M., Lilienfeld, S.O., & Nezworski, M.T. (2002). Effective use of projective techniques in clinical practice: Let the data help with selection and interpretation. *Professional Psychology: Research and Practice, 33,* 454-463. doi: 10.1037//0735-7028.33.5.454

Goh, D.S., Teslow, C.J., & Fuller, G.B. (1981). The practice of psychological assessment among school psychologists. *Professional Psychology, 12,* 696-706. doi: 10.1037/0735-7028.12.6.696

Goldberg, L.R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 2, pp.141-165). Beverly Hills, CA: Sage.

Green, P.E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *The Journal of Consumer Research, 5,* 102-123.

Green, P.E., & Srinivasan, V. (1990). Conjoint analysis in marketing: New developments with implications for research and practice. *The Journal of Marketing, 54,* 3-19.

Green, P.E., & Wind, Y. (1975). New way to measure consumers' judgments. *Harvard Business Review, 53,* 107-117.

Grove, W.M., Zald, D.H., Lebow, B.S., Snitz, B.E., & Nelson, C. (2000). Clinical versus mechanical prediction. *Psychological Assessment, 12,* 19-30. doi: 10.1037//1040-3590.12.1.19

Grove, W.M. (2005). Clinical versus statistical prediction: The contribution of Paul E. Meehl. *Journal of Clinical Psychology, 61,* 1233-1243. doi: 10.1002/jclp.20179

Halperin, J.M., & McKay, K.E. (1998). Psychological testing for child and adolescent psychiatrists: A review of the past 10 years. *Journal of the American Academy of Child and Adolescent Psychiatry, 37,* 575-584.

Hojnoski, R.L., Morrison, R., Brown, M., & Matthews, W.J. (2006). Projective test use among school psychologists: A survey and critique. *Journal of Psychoeducational Assessment, 24,* 145-159. doi: 10.1177/0734282906287828

Hosp, J.L., & Reschly, D.J. (2002). Regional differences in school psychology practice. *School Psychology Review, 31,* 11-29.

Huber, D.R. (2006). *Perceived self-efficacy and control for school psychologists.* (Doctoral dissertation). Retrieved from Dissertations & Theses: Full Text. (Publication No. AAT 3228563).

Hutton, J.B., Dubes, R., & Muir, S. (1992). Assessment practices of school psychologists: Ten years later. *School Psychology Review, 21,* 271-284.

Individuals with Disabilities Education Improvement Act, 20 U.S.C. § 1400 (2004). Regulations appear at 34 C.F. R. Part 300.

Jacob, S., & Hartshorne, T.S. (2007). *Ethics and law for school psychologists* (5th ed.). Hoboken, NJ: John Wiley & Sons, Inc.

John, O.P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin, & O. P. John (Eds.), Handbook of personality: Theory and research (2nd ed., pp. 102-138). New York: Guilford.

Kamphaus, R.W., & Frick, P.J. (2005). *Clinical Assessment of child and adolescent personality and behavior* (2nd ed.). New York, NY: Springer Science+Business Media, LLC.

Klein, D.N., Dougherty, L.R., & Olino, T.M. (2005). Toward guidelines for evidence-based assessment of depression in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34,* 412-432. doi: 10.1207/s15374424jccp3403_3

Kleinmuntz, B. (1967). *Personality measurement: An introduction.* Homewood, IL: Dorsey Press.

Kleinmutz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin, 107,* 296-310. doi: 10.1037/0033-2909.107.3.296

Knoff, H.M. (2002). Best practices in personality assessment. In A. Thomas & J. Grimes (Eds.). Best practices in school psychology IV (pp. 1281-1302). Bethesda, MD: National Association of School Psychologists.

Koonce, D.A. (2007). Attention deficit hyperactivity disorder assessment practices by practicing school psychologists. *Journal of Psychoeducational Assessment, 25,* 319-333. doi: 10.1177/0734282906298264

Koppitz, E.M. (1983). Projective drawings with children and adolescents. *School Psychology Review, 12,* 421-427.

Kratochwill, T.R., & Shernoff, E.S. (2003). Evidence-based practice: promoting evidence-based interventions in school psychology. *School Psychology Quarterly, 18,* 389-408.

Kuhfeld, W.F. (2005). *Marketing research methods in SAS: Experimental design, choice, conjoint, and graphical techniques.* Cary, NC: SAS Institute, Inc.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108,* 480-498. doi: 10.1037/0033-2909.108.3.480

Lachar, D., & LaCombe, J.A. (1983). Objective personality assessment: The personality inventory for children and its applications in the school setting. *School Psychology Review, 12,* 399-406.

Larry P. v. Riles, 343 F. Supp. 1306 (N.D. Cal. 1972); *aff'd*., 502 F.2d 963 (9th Cir. 1974); 495 F. Supp. 926 (N.D. Cal. 1979); *aff'd*., 793 F.2d 969 (9th Cir. 1984).

Levine, S. (2003). The self-deceptions and misconceptions of psychiatrists, psychologists, and other mental health professionals. *Adolescent Psychiatry, 27,* 325-343.

Lilienfeld, S.O., Wood, J.M., & Garb, H.N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1,* 27-66.

Louviere, J.J. (2000). *Stated choice methods: Analysis and applications.* Cambridge, UK: Cambridge University Press.

Maertz, C.P., Hassan, A., & Magnusson, P. (2009). When learning is not enough: A process model of expatriate adjustment as cultural cognitive dissonance reduction. *Organizational Behavior and Human Decision Processes, 108,* 66-78. doi: 10.1016/j.obhdp.2008.05.003

Maikovich, A.K. (2005). A new understanding of terrorism using cognitive dissonance principles. *Journal for the Theory of Social Behaviour, 35,* 373-397. doi: 10.1111/j.1468-5914.2005.00282.x

Matz, D.C., Hofstedt, P.M., & Wood, W. (2008). Extraversion as a moderator of the cognitive dissonance associated with disagreement. *Personality and Individual Differences, 45,* 401-405. doi: 10.1016/j.paid.2008.05.014

Matz, D.C., & Wood, W. (2005). Cognitive dissonance in groups: The consequences of disagreement. *Journal of Personality and Social Psychology, 88,* 22-37. doi: 10.1037/0022-3514.88.1.22

McClelland, D.C., Atkinson, J.W., Clark, R.A., & Lowell, E.L. (1953). *The achievement motive.* New York: Appleton-Century-Crofts.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*, 175-215.

McKimmie, B.M., Terry, D.J., Hogg, M.A., Manstead, A.S.R., Spears, R., & Doosje, B. (2003). I'm a hypocrite, but so is everyone else: Group support and the reduction of cognitive dissonance. *Group Dynamics: Theory, Research, and Practice, 7,* 214-224. doi: 10.1037/1089-2699.7.3.214

McMahon, R.J., & Frick, P.J. (2005). Evidence-based assessment of conduct problems in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34,* 477-505. doi: 10.1207/s15374424jccp3403_6

Medeiros, J.M. (2002). Barriers to compliance with evidence-based clinical practice guidelines. *The Journal of Manual & Manipulative Therapy, 10,* 7-9.

Miller, D.N., & Jome, L.M. (2008). School psychologists and the assessment of childhood internalizing disorders: Perceived knowledge, role preferences and training needs. *School Psychology International, 29,* 500-510. doi: 10.1177/0143034308096432

Miller, D.N., & Nickerson, A.B. (2006). Projective assessment and school psychology: Contemporary validity issues and implications for practice. *The California School Psychologist, 11,* 73-84.

Miller, D.N., & Nickerson, A.B. (2007). Projective techniques and the school-based assessment of childhood internalizing disorders: A critical analysis. *Journal of Projective Psychology and Mental Health, 14,* 48-58.

Mills, L.B., & Huebner, E.S. (1998). A prospective study of personality characteristics, occupational stressors, and burnout among school psychology practitioners. *Journal of School Psychology, 36,* 103-120. doi:10.1016/S0022-4405(97)00053-8

Mullen, E., & Skitka, L.J. (2006). Exploring the psychological underpinnings of the moral mandate effect: Motivated reasoning, group differentiation, or anger? *Journal of Personality and Social Psychology, 90,* 629-643. doi: 10.1037/0022-3514.90.4.629

Murray, H.A. (1943). *Thematic apperception test.* Cambridge, MA: Harvard University Press.

Nagy, T.F. (2005). *Ethics in plain English: An illustrative casebook for psychologists (*2nd ed.). Washington, DC: American Psychological Association.

National Association of School Psychologists. (n.d.). What is a school psychologist? Retrieved from http://www.nasponline.org/about_sp/whatis.aspx

National Association of School Psychologists. (2000). Professional conduct manual, principles for professional ethics, and guidelines for the provision of school psychological services. Bethesda, MD: Author.

National Association of School Psychologists. (2003). School psychologists' involvement in the role of assessment (Position Statement). Bethesda, MD: Author.

Olsen, R.A. (2008). Cognitive dissonance: The problem facing behavioral finance. *Journal of Behavioral Finance, 9,* 1-4. doi: 10.1080/15427560801896552

Oskamp, S. (1984). Theories in Applied Social Psychology: Appendix. In S. Oskamp, *Applied Social Psychology,* 392-409.

Ozonoff, S., Goodlin-Jones, B.L., & Solomon, M. (2005). Evidence-based assessment of autism spectrum disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34,* 523-540.

Pelham, Jr., W.E., Fabiano, G.A., Massetti, G.M. (2005). Evidence-based assessment of attention deficit hyperactivity disorder in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34,* 449-476. doi: 10.1207/s15374424jccp3403_5

Peretti-Watel, P. (2006). Cognitive dissonance and risk denial: The case of cannabis use in adolescents. *The Journal of Socio-Economics, 35,* 1032-1049. doi: 10.1016/j.socec.2005.11.023

Peterson, D.W., & Batsche, G.M. (1983). School psychology and projective assessment: A growing incompatibility. *School Psychology Review, 12,* 440-445.

Prout, H.T. (1983). School psychologists and social-emotional assessment techniques: Patterns in training and use. *School Psychology Review, 12,* 377-383.

Pruitt, J.A., Smith, M.C., Thelen, M.H., & Lubin, B. (1985). Attitudes of academic clinical psychologists toward projective techniques: 1968–1983. *Professional Psychology: Research and Practice, 16,* 781-788. doi: 10.1037/0735-7028.16.6.781

Rehabilitation Act of 1973, 29 U.S.C. § 794, (1973). Regulations appear at 34 C.F.R. Part 104 (1996).

Reschly, D.J. (2000). The present and future status of school psychology in the United States. *School Psychology Review, 29,* 507-522.

Reynolds, C.R., & Kamphaus, R.W. (2004). *Behavior assessment system for children* (2nd ed.). Circle Pines, MN: American Guidance Service.

Rodriguez, R., Marchand, E., Ng, J., & Stice, E. (2008). Effects of a cognitive dissonance-based eating disorder prevention program are similar for Asian American, Hispanic, and White participants. *International Journal of Eating Disorders, 41,* 618-625. doi: 10.1002/eat.20532

Romerdahl, E. (2009). The shame of the legal profession: Why eighty percent of those in need of civil legal assistance do not receive it and what we should do about it. *Georgetown Journal of Legal Ethics, 22,* 1115-1134.

Rossini, E.D., & Moretti, R.J. (1997). Thematic Apperception Test (TAT): Practice recommendations from survey of clinical psychology doctoral programs accredited by the American Psychological Association. *Professional Psychology: Research and Practice, 28,* 393-398. doi: 10.1037/0735-7028.28.4.393

Salkind, N.J. (2006). *Tests and measurement for people who think they hate tests and measurement*. Thousand Oaks, CA: Sage Publications, Inc.

Salvia, J., Ysseldyke, J.E., & Bolt, S. (2007). *Assessment in special and inclusive education* (10th ed.). Boston, MA: Houghton-Mifflin.

Sattler, J.M. (2008). *Assessment of children: Cognitive applications* (5th ed.). San Diego, CA: Sattler.

Sattler, J.M., & Hoge, R.D. (2006). *Assessment of children: Behavioral, social, and clinical foundations* (5th ed.). San Diego, CA: Sattler.

Shapiro, E.S., & Heick, P.F. (2004). School psychologist assessment practices in the evaluation of students referred for social/behavioral/emotional problems. *Psychology in the Schools, 41,* 551-561. doi: 10.1002/pits.10176

Sherer, M., Maddux, J., Mercandante, M., Prentice-Dunn, S., Jacobs, B., & Rogers, R.W. (1982). The self-efficacy scale: Construction and validation. *Psychological Reports, 51,* 663-671.

Silverman, W.K., & Ollendick, T.H. (2005). Evidence-based assessment of anxiety and its disorders in children and adolescents. *Journal of Clinical Child & Adolescent Psychology, 34,* 380-411. doi: 10.1207/s15374424jccp3403_2

Smith, D. (1984). Practicing school psychologists: Their characteristics, activities, and populations served. *Professional Psychology: Research and Practice, 15,* 798-810. doi: 10.1037/0735-7028.15.6.798

Starzyk, K.B., Fabrigar, L.R., Soryal, A.S., & Fanning, J.J. (2009). A painful reminder: The role of level and salience of attitude importance in cognitive dissonance. *Personality and Social Psychology Bulletin, 25,* 126-137. doi: 10.1177/0146167208325613

Stinnett, T.A., Havey, J.M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. *Journal of Psychoeducational Assessment, 12*, 331-350. doi: 10.1177/073428299401200403

Sweeney, J.A., Clarkin, J.F., & Fitzgibbon, M.L. (1987). Current practice of psychological assessment. *Professional Psychology: Research and Practice, 18,* 377-380. doi: 10.1037/0735-7028.18.4.377

Tharinger, D.J., Pryzwansky, W.B., & Miller, J.A. (2008). School psychology: A specialty of professional psychology with distinct competencies and complexities. *Professional Psychology: Research and Practice, 39,* 529-536. doi: 10.1037/0735-7028.39.5.529

Thelen, M.H., Varble, D.L., & Johnson, J. (1968). Attitudes of academic clinical psychologists toward projective techniques. *American Psychologist, 23,* 517-521. doi: 10.1037/h0026195

Toomey, K.D., Levinson, E.M., & Morrison, T.J. (2008). The vocational personality of school psychologists in the United States. *School Psychology International, 29,* 418-425. doi: 10.1177/0143034308096436

Tupes, E.C., & Christal, R.E. (1992). Recurrent personality factors based on trait ratings. *Journal of Personality, 60,* 225-251.

T.Y. v. New York City Department of Education, Region 4, 584 F.3d 412 (2d Cir. 2009).

VanZandt, C.E. (1990). Professionalism: A matter of personal initiative. *Journal of Counseling & Development, 68,* 243-245.

Vukovich, D.H. (1983). The use of projective assessment by school psychologists. *School Psychology Review, 12,* 358-364.

Watkins, C.E., Campbell, V.L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice, 26,* 54-60. doi: 10.1037/0735-7028.26.1.54

Wechsler, D. (2003a). *WISC-IV administrative and scoring manual*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003b). *WISC-IV technical and interpretive manual*. San Antonio, TX: The Psychological Corporation.

Wenger, R.D., & Pryzwansky, W.B. (1987). Implementation status of the APA guidelines for delivery of services by school psychologists. *Professional Psychology: Research and Practice, 18,* 461-467. doi: 10.1037/0735-7028.18.5.461

Whitney, S.R., & Tryon, G.S. (2009). Assessment practices of school psychologists. Unpublished Manuscript, City University of New York: Graduate School and University Center.

Wilson, M.S., & Reschly, D.J. (1996). Assessment in school psychology training and practice. *School Psychology Review, 25,* 9-23.

Youngstrom, E.A., Findling, R.L., Kogos Youngstrom, J., Calabrese, J.R. (2005). Toward an evidence-based assessment of pediatric bipolar disorder. *Journal of Clinical Child and Adolescent Psychology, 34,* 433-448. doi: 10.1207/s15374424jccp3403_4