RELIABILITY AND VALIDITY OF CAREER ASSESSMENT INSTRUMENTATION

Approved Richard N Rogers Date 2/3/10

RELIABILITY AND VALIDITY OF CAREER ASSESSMENT INSTRUMENTATION

A Seminar Paper Research Proposal

Presented to

The Graduate Faculty

University of Wisconsin-Platteville

In Partial Fulfillment of the

Requirement for the Degree

Master of Science

In

Adult Education

By

Heidi Oniszczuk

2010

Abstract

RELIABILITY AND VALIDITY OF CAREER ASSESSMENT INSTRUMENTATION

Heidi Oniszczuk

A review of the literature was done to assess validity and reliability of selected career assessment instrumentation. The career assessment tests that were chosen are Myers-Briggs Type Indicator, Strong Interest Inventory, True Colors, and the Clifton StrengthsFinder. A few different kinds of reliability and validity were looked at.

Test- retest reliability was the most popular means of assessing reliability, but split half reliability measures were also used. Many kinds of validity were assessed. They include, content validity, concurrent validity, construct validity, criterion validity, and predictive validity.

With a few exceptions with some aspects of the instruments, each career assessment demonstrated the various kinds of reliability and validity.

TABLE OF CONTENTS

Chapter I

Introduction

Frank Parsons (1854-1908), is the father of vocational guidance. He was the first to give credence to the idea that different people are suited for different careers. Parsons indicated that one needed a clear understanding of him/herself, their personality, abilities, and interests before deciding on a career path. To illustrate how important this process of vocational guidance is, Parsons gives an example:

> No one would think of building a dwelling or a business block without carefully selecting an appropriate and advantageous site and drawing a well-considered plan with the help of an architect or expert builder. And in building a career it is quite as important to make a wise location, lay the foundation properly, and work up by a well-considered, scientific plan (Parsons, 1909 p.101)

Thus vocational guidance was born, and shortly thereafter, career assessments were being introduced. Many assessment tests were developed and continue to be developed today. The Strong Interest Inventory was developed in 1927, and the Myers-Briggs Type Indicator was developed in the 1940's. The first time the United States Government needed career assessment tests was during World War one. Military psychologists had to figure out the best place for each man to work. They needed an objective way to figure out who would be best on the front lines of battle and which men would be better as cooks.

Today, career assessments are still being used and developed. The True Colors was developed in 1978, and the Clifton StrengthsFinder in the 1990's. Myers, McCaulley, Quenk, and Hammer outline the advantages of the use of assessment tests:

1. Tests are more objective than observations.
2. Tests take less time than interviewing to get the necessary information.
3. Tests provide norms and therefore a standard of comparison for the client to evaluate

himself. Tests can reveal predispositions that might not surface in interviews or be recognized in the client.

4. Tests provide clients with information about themselves and their relationship to the working world (1998, p. 3).

Students looking for assistance with their career paths are often called 'undecided'. However, Hawkins-Wilding, Director of Advising and Career Services at the University of Wisconsin- Parkside in Kenosha, Wisconsin says that, "Undecided seems to give a negative connotation to the student". She prefers to give them a positive, proactive title, "exploring students" (S. Hawkins- Wilding, personal communication, July, 6, 2009). The title "exploring students" gives credence to the importance of getting wise counsel and expert advice as Parsons indicated.

Career counseling takes place in high schools, colleges and job centers. A student looking for guidance will talk over concerns with a counselor and from that conversation the counselor may suggest that the student take a career assessment test. Assessments are particularly useful for students who wish to "investigate majors, explore careers, change careers, pursue graduate school, identify co-curricular activities, and engage in lifelong learning endeavors" (M. Smith, personal communication, July 12, 2009).

It is not possible to research all the tests and assessments since there are a lot of them. This paper is focused on four career assessment tests that are currently used at University of Wisconsin Parkside in Kenosha, Wisconsin. Those tests are: Myers-Briggs Type Indicator (MBTI); Strong Interest Indicator (SII); True Colors; and the Clifton StrengthsFinder (CSF).

Statement of the Problem

Career assessment tests are many and varied. The different kinds of tests help students

determine varying information such as their individual aptitudes, strengths, personalities and talents. Career counselors need to become familiar with all the tests to better serve exploring students. Not only do career counselors need to know the purpose of each test, but they need to understand how reliable and valid each test is with each type of student. The purpose of this paper is to explain the abilities of each test, and review the literature to discover the reliability and validity of each test. Reliability and validity of each test are also researched with use for varying types of ethnicities and races.

Delimitations of the Research

The research was conducted through the Karmann libraries (University of Wisconsin-Platteville), Google searches through the World Wide Web, and interviews of personnel in the career center at University of Wisconsin- Parkside over a period of seven (7) months. Primary searches were conducted via Internet through Sage Journals Online and EBSCO Host via ERIC. Key search topics will include "career instrumentation," "reliability/validity and True Colors," "reliability/validity and Myers Briggs Type Indicator," "reliability/validity and Clifton StrengthsFinder," and "reliability/validity and Strong Interest Inventory."

Method of Approach

A review of literature was done on topics related to history, research, and experiments of the reliability and validity of career assessment tests. A majority of the history, research and experiments came from manuals. The Myers-Briggs Type Inventory and the Strong Interest Inventory both had manuals filled with research. No such manuals were found for the True Colors and Clifton StrengthsFinder assessments. The Clifton StrengthFinder came from the Gallup Organization which performed extensive research to validate this test. However, no published articles were found for the True Colors assessment test. The research used for this

paper was found on the True Colors website from students earning either a masters or doctorate degree.

Searches through Sage Journals Online provided most of the supplemental information. These studies provided additional experiments on the reliability and validity of the tests with cross- cultural participants and various abilities that the tests are not currently known for, such as the ability of the tests to predict college major. The findings were summarized and recommendations made.

Definition of Terms

Reliability: Is the extent to which a test yields consistent results. (Reliability and Validity, 2009).

Test-Retest Reliability: Measures a tests consistency over time. A second test is given after a certain amount of time usually three to six months. The results from the first test and second test are compared to check for consistency and given a correlation statistic measurement between 0.00 (no consistency) to 1.0 (perfect consistency).

Split – Half Reliability: Checks reliability of a test in one day. One person is given a test, then the researcher splits the test in two halves - usually odd and even numbered items. The scores are then measured for consistency.

Validity: Is the extent to which a test measures or predicts what it is supposed to (Reliability and Validity, 2009).

Predictive Validity: Is when the researcher waits over time to see if the test was accurate in predicting future outcomes.

Concurrent Validity: Compares performance on the test with another already validated test.

Criterion Validity: Compares test results with the judgments of a panel of clinical psychologists.

Construct Validity: Compares the hypothetical basis of the test against a measurable basis. For example, introverts tend to have a lack of comfort with their environment and this can be measured as anxiety, hesitating, and yielding.

Content Validity: Is a non-statistical type of validity that involves an examination of the test content to determine whether it covers a representative sample of the domain to be measured

Chapter II

Review of the Literature

Career assessment tests are a useful tool in career counseling. They help students find out more about their interests, personalities, strengths and aptitudes. These tests help counselors guide the counseling session. This review will focus on the following assessment tests: Meyers-Briggs Type Indicator, Strongs Interest Inventory, True Colors, and Clifton StrengthsFinder.

Meyers-Briggs Type Indicator

The Myers-Briggs Type Indicator (MBTI) was developed by Katharine Cook Briggs and her daughter Isabel Briggs Myers in the 1940's. It is a self-report instrument that helps to identify an individual's strengths and personality preferences. Briggs and Myers wanted to match people (and their personalities) to the right kind of job. They based their test on Carl Jung's theories of psychological type preferences.

To measure differences in personality, respondents choose between preferences on a self-report test. These preferences are named: Extraversion- Introversion (E-I); Sensing- Intuition (S-N); Thinking- Feeling (T-F); and Judging- Perceiving (J-P). According to the Myers-Briggs foundation, someone with a personality type of extraversion may favor a career with more human interaction, a salesman for example, while an introverted person may enjoy working with numbers like an accountant. Because of these and other personality differences in students, it is important to know and understand ones' personality in order to make better career choices.

Studies testing the reliability and validity of the MBTI became plentiful since the early 1990's. One reason might be that despite the test's popularity, some skeptics were coming on the scene. One such skeptic was Druckman and Bjork who wrote in their book, *In the Minds Eye: Enhancing Human Performance,* "the popularity of this instrument (MBTI) in the absence of

proven scientific worth is troublesome" (as reported in Myers p. 435). However, many studies now prove the reliability and validity of the MBTI.

Myers, McCaulley, Quenk, and Hammer (Myers et al.) authored the MBTI Manual in 1998. They did reliability and validity tests on Form "M" of the MBTI. The scores came out favorably for the MBTI, however the authors give a list of factors that can affect reliability of the MBTI. Such factors are:

1. The ability and understanding of the student.
2. Higher achieving students have higher reliabilities scores than lower achieving students.
3. Students in college preparatory classes and advanced placement classes have higher reliabilities than other groups of students.
4. Students with higher intelligence scores and students in college have higher reliability scores than students with lower intelligence scores and students in high school (p. 159).

Myers et al. (1998) did split – half reliability tests on Form M of the MBTI. They split the test in half to compare the scores a few different ways. First, they did a logical split-half of E-I, S-N, T-F and J-P ranged from .90 - .92. Next, they did a consecutive split–half of E-I, S-N, T-F and J-P. Scores ranged from .89 - .94. Finally, they split the test by item format. This means they split the test into word pairs and phrase questions. The reliability scores ranged from .90 - 93.

Next, Myers et al. (1998) looked at reliability scores among ethnic groups of college students. Reliabilities scores for African– American's ranged from .80 - .91. American– Indian scores ranged from .86 - .96. Asian or Pacific Islander's reliability scores ranged from .82 - .91. Latino/ Latina/ Hispanic scores ranged from .84 - .90. Among adult ethnic groups, African–

American's reliabilities ranged from .84 - .91 and Latino/ Latina/ Hispanic adults, scores ranged from .87 - .91.

Myers et al. (1998) also did test – retest reliabilities on Form "M" over a four week period. The scores ranged from .83 - .97. When students report a change in their type (E-I, S-N, T-F and J-P) Myers et al. say it is most likely to occur in scales where the original preference clarity was low. The reliability coefficient for T-F remains the lowest of the four scales. They ranged from .83 - .94. The range for the rest of the scales are: J-P - .90- .95; S-N - .89- .97; and E-I- .93- .95.

For validity (the extent to which a test measures or predicts what it is supposed to) of the MBTI, Myers et al. (1998) confirmed construct validity through a confirmatory and exploratory factor analyses. An extrovert is defined as people who like action in their environment. For validating construct validity, Myers et al. measured extroversion by dominance, assertiveness, social boldness, capacity for status, leadership, change, aggression, and exhibition. Through the analyses, the four factor model scored .949 for adjusted goodness of fit, .967 for the non-normed fit index and -.008 for the median of the fitted residuals. Myers et al. concludes that "these results indicate an excellent fit" (1998 p. 173) of the four factor model E-I, S-N, T-F and J-P.

Myers et al. (1998) also researched to confirm concurrent validity. They point to the correlations of the four preference scales E-I, S-N, T-F and J-P with other scales. They compare the MBTI with several other scales such as the Jungian Type Survey, Millon Index of Personality Styles, Internality Scale from the California Psychological Inventory, and The Sixteen Personality Factors Questionnaire (NEO-PI-). They conclude that these comparisons "support the predictions of type theory regarding the meaning of and the behaviors believed to be associated with the four dichotomies" (1998 p. 173).

Other tests have been done on the MBTI. Pulver and Kelly (2008) did a study to see if the MBTI could predict academic major selection of undecided university students. They found that the MBTI is not a good instrument for predicting what undecided students choose for a college major. The Myers-Brigg's Foundation recommends that the MBTI be used to help in deciding what specific areas of various fields of study a person prefers such as psychology, law, medicine, education or business. The MBTI cannot detect which field of study to take, but can help in deciding where to specify in any chosen field; for example, if a person decides to go into business, the MBTI can help them decide the best area in business to go into such as: secretary, manager, chief executive officer or sales. Other career assessment tools, such as the Strong's Interest Inventory, was developed to help exploring students decide which major to take.

Strong Interest Inventory

The Strong Interest Inventory (SII) was developed by Strong in 1927. Strong's research focused on how satisfied employees in a specific occupation report similar interests. Strong hypothesized that he could differentiate these interests among careers and developed an inventory test to help exploring students make a decision about their career paths. The Strong Interest Inventory is for students age 15 and older and provides information on 109 different occupations, including those that require vocational training, technical training and professional education.

The SII is based on the idea that people are more satisfied and productive when they work at jobs they find interesting and when they work with people whose interests are similar to their own. Items measure interest in a variety of occupations, occupational activities, hobbies, leisure activities, and types of people. The SII also matches a person's interest with various careers and occupations.

The SII measures students' interests and compares them to those of people already working in a wide array of occupations. The test has 317 questions. Students rate each question with one of three choices: like, indifferent, or dislike. The SII not only assesses patterns of interest, but also occupational themes. The General Occupational Themes (GOT) are: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional (RIASEC). First, this paper will reveal the reliability and validity coefficients of the GOT's.

In 1994, Harmon, Hansen, Borgan, and Hammer, (Harman et al.) did a test-retest reliability test in four samples (A-D) for the GOT's:

Sample A: 191 nation-wide sample of employed adults in fifteen diverse occupations. 110 were women and 81 were men. They were retested after three to six months.

Sample B: 84 college students, 54 women and 30 men, in a psychological measurement course retested after one month. They were students at Iowa State University enrolled in an upper level graduate course, consisting of mainly junior and senior psychology majors.

Sample C: 79 college students, 56 women and 23 men, in a career development course at the University of North Carolina at Greensboro were retested after three months. Their ages ranged from 18-25 years with one thirty year old. These students were sophomores, juniors or seniors.

Sample D: 87 college students, 61 women, 26 men, in a career development course at the University of Illinois were retested after three months. This group consisted of freshman through seniors.

Table 1

Results of the test- retest reliabilities for each sample (A-D) with each GOT (RIASEC) theme.

	Realistic	Investigative	Artistic	Social	Enterprising	Conventional
Sample A	.92	.91	.91	.85	.84	.87
Sample B	.88	.84	.86	.84	.86	.86
Sample C	.78	.85	.91	.89	.77	.78
Sample D	.84	.86	.86	.82	.74	.82

The test-retest reliabilities in sample A was the highest among the samples. Realistic: .92, Investigative: .91, and Artistic: .91 had the highest reliability. This implication for counseling is that these preferences among students are likely to be stable over time. These high test- retest reliability studies indicate that the use of GOT's in career planning is advantageous. Hansen and Campbell in 1985 did a test- retest study over a three year period of the GOT's. Results were still showing high reliability. The average score among the themes was .81 with a range of .78- .82 (as cited in Harmon et al. 1994 p. 58).

The GOT's went under concurrent validity measures when Hansen and Campbell compared the results with another similar inventory in 1985 called the Holland's Vocational Preference Inventory. Correlations between the tests were high: .765. This indicates that the two inventories measure similar interest dimensions (as cited in Harmon et al. 1994 p.59).

Basic Interest Scales (BIS) is another aspect of SII. BIS's are subdivisions of the GOT's. GOT's are more diverse in content span and BIS's are more focused on a specific interest domain. For example, the "R" GOT include BIS's of agriculture, nature, athletics, military activities, and mechanical activities. Additionally, each BIS goes into further focus. For example, the mechanical activities scale includes the following items: auto mechanic, mechanical engineer, operating machinery, solving mechanical puzzles, and having mechanical ingenuity/inventiveness.

Using the BIS in counseling helps students understand the underlying interests measured

by the GOT's. Each BIS question is weighted +1 for a "like" response, 0 for an "indifferent" response, and a -1 for a "dislike" response. Clients who have a high number of "like" responses (60%- 70% or higher) would be said to have "high interest" or "very high interest" in that area. In counseling the top three to four areas and the lowest areas should be considered.

Table 2

Results of the test- retest reliability range of scores for each sample (A-D) of the BIS's portion of the SII.

	Sample A	Sample B	Sample C	Sample D
Retested after:	3-6 months	1 month	3 months	3 months
Range:	.80-.92	.78-.93	.66-.91	.72-.90

Sample A, the working employed adults continues to have the highest reliability.

The BIS's have content validity. Each scale is focused on one content area and the items reflect this focus. For example, the Science scale contains items like astronomer, biologist, chemist, and working in a research laboratory. Harmon et al. concludes that these items have content reliability because responses to them provide direct information about the person's feelings toward scientific activities. The same is true of the other scales (1994, p. 85).

For concurrent and construct validity, Harmon et al. researched scores of people who are currently in occupations. The researchers found the BIS's to have "substantial validity" in that an artist scored high on the art scale, scientists scored high on the science scale and teachers scored high on the teaching scale (1994, p. 91). Also Harmon et al. found that they scored low or average on scales that did not relate to their occupations (1994, p. 91).

Hansen and Campbell (1985) found that the BIS's have predictive validity (as cited in Harmon et al, 1994, p 91). The researchers found that clients who scored "high" or "very high" in a specific BIS such as science, were found to be, years later, in an occupation that was

scientific in nature. This research by Hansen and Campbell was done on the previous edition of the SII, but since they BIS's did not change very much, they generalized their findings to the 1994 SII.

Among other cultures, Flores, Spanierman, Armstrong, and Valez (Flores et al.) concludes that "career counselors can continue to use the SII with Mexican American youth to assess career interests and self- efficacy" (2006, p. 197). However, Flores et al. (2006) recommends replicated research in regard to test-retest reliability and validity among Mexican American youth. Betz and Gwilliam (2002) report that the SII is "adequately reliable" among African American and European American students, (as cited in Flores et al., 2006, p. 186). Further, Harmon et al., (1994) concludes that various racial and ethnic group scores have not differed significantly, only slightly, from the normative samples. In 2002, Fouad researched various racial and ethnic groups on the SII and found that there were greater differences within than between racial and ethnic groups (as cited in Flores et al., 2006, p. 186).

True Colors Word Cluster

True Colors was developed by Lowry in 1978. It is a self-report, Likert-type scale instrument that utilizes color as a metaphor to describe personality. The colors chosen to represent each personality type include Green, Blue, Orange, and Gold. Each color clusters unique personality characteristics. These characteristics help students understand the best type of career path for their unique personality style.

To test reliability of the True Colors test, a test- retest format was used. Participants were enrolled in one of four classes: a True Colors Awareness workshop, a Level One Certification Training, or one of two, level one-hundred Psychology college courses. In total, the participants numbered 167. The participants had a wide range of occupations, and educational backgrounds.

Among all these participants the mean reliability coefficient was .940 (Whichard, 2006). The time between test and re-test ranged from thirty-six days to forty- two days.

A study conducted by Honaker (2003) revealed support for content validity between the True Colors Personality Inventory and the Meyers-Briggs Type Indicator. This study utilized data from fifty-six graduate students (38 females and 18 males) enrolled in a graduate level career development course. All students completed the assessments during the fall 1998 and spring 1999 semesters. The study revealed significant positive correlations between the two tests in terms of personality types and characteristics. The correlations are as follows:

Table 3

Results of content validity between True Colors Themes and MBTI preferences

MBTI:	Perceptive	Judging	Thinking	Feeling
True Colors:	Orange	Gold	Green	Blue

Whichard (2006) took Honaker's research a step further and examined the relationship between the True Colors and other tests (MBTI and DISC) to see if True Colors test has concurrent validity. Whichard's (2006) study involved sixty-seven participants from three Awareness Workshops, and two, one-hundred level psychology courses. Whichard (2006) was looking for the extent to which True Colors compared with MBTI and True Colors compared with DISC results indicated similar personality, psychological, behavioral and temperament characteristics. The following chart reveals the correlations between True Colors and MBTI.

Table 4

Results of concurrent validity between True Colors and MBTI

MBTI	Perceptive	Judging	Thinking	Feeling
True Colors	Orange	Gold	Green	Blue
Correlation Coefficient	.751	.776	.861	.834

The strength of the correlations between the MBTI and True Colors is high, indicating concurrent validity. The following chart (Table 5) reveals the correlations between the True Colors and the DISC.

Table 5

Results of concurrent validity between True Colors and DISC

True Colors	Orange	Gold	Green	Blue
DISC	Dominance	Steadiness	Compliance	Influencing
Correlation Coefficient	.734	.807	.787	.819

Again, the strength of the relationship between the DISC and True Colors indicates that True Colors has concurrent validity.

Whichard (2006) also tested construct validity with the True Colors. Five groups, with a total of 117 participants, were asked to rate the True Colors assessment in its ability to accurately predict ten theoretical concepts. The five groups were participants involved in either a True Colors Awareness workshops or a Level One Certification Training. The ten theoretical concepts are as follows: Communication Preferences, Learning Preferences, Values, Strengths, Needs, Stressors, Diversity, Workplace compatibility, Self-esteem, and Introversion and Extroversion. The following chart (Table 6) illustrates the average score the five groups rated the True Colors as having the ability to predict the ten theoretical concepts.

Table 6

Results of construct validity of the True Colors

Theoretical Concept:	Group Ratings (6=highly accurate to 1= no accuracy)
	Mean:
Communication Preferences	5.18

Learning Preferences	5.30
Values	5.22
Strengths	5.28
Needs	5.06
Stressors	4.96
Diversity	5.34
Workplace Compatibility	4.96
Self-esteem	5.52
Introversion/Extraversion	3.18

All ratings were high except Introversion/Extraversion. Thus, nine out of ten theoretical concepts have construct validity.

Clifton StrengthsFinder

The Clifton StrengthsFinder (CSF) was developed from the research of Clifton. Clifton is considered the father of strengths-based psychology and grandfather of positive psychology. Through Clifton's research, he has discovered two basic principles: "1) Each person's talents are enduring and unique and 2) each person's greatest room for growth is in the areas of his or her greatest strength" (Buckingham and Clifton, 2001, p. 8). The CSF was devised to help career counselors incorporate strengths into the planning of students' career paths, and to help students discover they have strengths, figure out what those strengths are and how to incorporate them into a career. If students are able to find a career that utilizes their individual strengths, according to Buckingham and Clifton, they will be "more productive, more fulfilled and more successful" (Buckingham and Clifton, 2001, p. 5).

To test reliability and validity, Shreiner (2006) tested students from fourteen colleges and universities (five community colleges and nine universities). 750 students were involved from the start, but only 479 completed all phases of the study, and of these, only 438 had usable test results. Only data from the 438 students who had successfully completed all phases of the study were used. Of the 438 students, 54% were female, 46% were male, 46% were first year students,

31.5% were second year students, 8.7% were third year students and 10.8% were fourth year students. Three percent did not report their class level. This sample was representative of the students attending the colleges and universities. The ethnic percents are as follows: Caucasian 76%, Asian 13.6%, Hispanic 5%, African-American 4.3% and multiethnic 1.2%.

The CSF is composed of thirty-four themes. These themes describe talents and each participant is given their top five themes as a result of the CSF. Test-retest reliability tests were given eight to twelve weeks apart, and students were not given their test results after the first test. The mean test-retest reliability estimate was .70 (Shreiner, 2006, p.5). Each theme's reliability estimates, along with gender differences and racial differences are included on the following chart (Table 7).

Table 7

Results of the test-retest reliability estimate mean among females, males, Caucasian's and minority students

CSF Theme	Test-Retest Reliability Estimate	Females (n=228)	Males (n=180)	Caucasian Students	Minority Students
Achiever	.78	.78	.78	.75	.81
Activator	.52	.50	.58	.52	.39
Adaptability	.69	.69	.70	.68	.65
Analytical	.76	.76	.77	.74	.81
Arranger	.63	.60	.67	.61	.71
Belief	.77	.76	.77	.79	.74
Command	.67	.66	.71	.69	.59
Communication	.75	.72	.78	.77	.66
Competition	.80	.80	.80	.79	.73
Connectedness	.75	.68	.78	.75	.75
Consistency	.53	.51	.50	.51	.63
Context	.65	.61	.73	.65	.64
Deliberative	.81	.84	.80	.85	.70
Developer	.79	.75	.83	.78	.79
Discipline	.84	.84	.86	.82	.90

Empathy	.74	.72	.69	.75	.74
Focus	.68	.70	.62	.69	.73
Futuristic	.69	.72	.66	.70	.65
Harmony	.62	.64	.65	.57	.75
Ideation	.65	.67	.67	.70	.53
Includer	.66	.66	.66	.66	.66
Individualization	.60	.57	.62	.60	.55
Input	.77	.78	.76	.76	.71
Intellection	.80	.84	.75	.84	.63
Learner	.78	.80	.80	.80	.76
Maximizer	.55	.51	.61	.52	.55
Positivity	.80	.80	.81	.78	.83
Relator	.65	.64	.61	.68	.53
Responsibility	.70	.66	.77	.69	.75
Restorative	.70	.68	.79	.68	.72
Self-Assurance	.65	.58	.78	.68	.52
Significance	.65	.65	.65	.67	.52
Strategic	.65	.67	.69	.70	.62
Woo	.78	.79	.81	.76	.82

(Schreiner, 2006, p. 10-12).

The themes with the highest reliability ratings have coefficients of .80 or higher. These themes are Competition, Deliberative, Discipline, Intellection, and Positivity. The lowest coefficient ratings have estimates of .60 and lower. These themes are Activator, Consistency, Indivdiualization, and Maximizer. In examining gender differences on the CSS, Schreiner found some differences between men and women. Women scored higher than men on the themes of Achiever, Belief, consistency, Developer, Discipline, Empathy, Harmony, Input, and Responsibility. Men scored higher than women on the theme of Ideation. In examining racial differences, Schreiner says that the differences on the CSF were slight (2006, p.8). However, ethnic minorities scored higher than Caucasian students on the themes of Significance, Harmony, and Analytical. Caucasian's scored higher than ethnic minorities on the themes of Adaptability, Self-Assurance, and Strategic.

Schreiner did additional research on the themes over time and evaluated how consistent students' top five themes were. The possibility of retaining the exact same themes is small due to

the fact that there are 278,256 possible combinations of top five themes. Despite this fact, 52% of the students had three or more themes that remained in their top five. Another 35% retained two of their themes over time, 11% retained one of their themes over time, and 2% did not retain any of the themes (Schreiner, 2006, p. 6).

Concurrent validity was obtained in Schreiner's study by correlating student's scores with two other validated assessment tests: the CPI-260 and the 16PF. 137 predicted relationships between CSF themes and their counterparts on the other tests were explored. A total of 128 or 93.4% of these predictions were confirmed by significant correlation coefficients (Schreiner, 2006, p. 7).

Next, Schreiner examined construct validity. This was done by assessing how each item fit into a particular theme. To evaluate this, a hierarchical cluster analysis was performed. Sireci (2001) says that a 70% pairwise hierarchical cluster indicates adequate construct validity (as cited in Schreiner, 2006, p. 8). 95% of the themes pairs met Sireci's recommendation. The average percent was 90% (Schreiner, 2006, p. 8).

Chapter III

Conclusions and Recommendations

A majority of the history, research and experiments came from manuals on the Myers-Briggs and the Strong Interest Inventory. No such manuals were found for the True Colors and Clifton StrengthsFinder assessments. The CSF came from the Gallup Organization which performed extensive research to validate this test. However, no published articles were found for the True Colors assessment test. The research used for this paper was found on the True Colors website from students earning either a masters or doctorate degree. Because there was a lot of research on the MBTI, SII, and the CSF, the findings seem all the more validated. However, since there was little research done on the True Colors, it is recommended that the research be replicated to validate the initial findings.

To determine how reliable and valid each test is, organizations offer some statistics to interpret correlation coefficients as significant or not. Wilson outlined definitions of small, medium and large correlations. They are as follows:

Small: 0.10

Medium: .25

Large: .40 (p. 1).

However, the American educational Research Association, the American Psychological Association, National Council on Measurement in Education agree that coefficient of .70 is needed for determining reliability (as sited in Schreiner, 2006, p.5). Most aspects of each test passed the APA, NCME standards for reliability, and all aspects of each test passed Wilson's standards.

This means that each assessment, MBTI, SII, CSF and True Colors, are all valid and reliable. Perhaps one should not be surprised to find that the career assessment tests used in career counseling should have high reliability and validity. In speaking with the Director of the Career Center at University of Wisconsin Parkside, Kenosha, Wisconsin, Ms. Hawkins- Wilding reports that, "if they are not reliable then we should not be using them!" (Personal communication, July 6, 2009).

Career counselors need to make sure they know what each test is designed to do. It is a mistake to assume that a career assessment test can do everything. Career counselors need to understand the limits of each assessment for the best results. One need to pay attention to age limits. The MBTI and the SII are for students over the age of 15 and the MBTI specifically said that reliability was lower for underachieving students. The Myers & Briggs Foundation suggests that when using the MBTI with underachieving students, that the results can be used for a starting point in discussing career options.

In locating research for this study, there was only one longitudinal study. Most test-retest reliability studies waited six month between retesting. In this respect, it would be advantageous to see correlation coefficients after waiting three years and compare them to the correlation coefficient after waiting six months.

Colleges and universities are filled with students from varying ethnicities and races. Since every student and his/her experience is unique, it is imperative that career counselors understand how to counsel each student. The research showed that various ethnicities and races can take each of the assessments and rely on reliable and valid results. However, career counselor needs to be familiar with the research to make sure that their student's race or ethnicity was represented in the research for each aspect of the assessment tests. The ethnicities and races

that were represented in the tests are as follows.

Table 8

The reliability scores (when available) of ethnicities and races represented in each test

Reliability Scores	African Americans	Hispanic/ Latino/ Latina	Asian or Pacific Islander	American-Indians	Multi-ethnic
MBTI	.80-.91	.84-.90	.82-.91	.86-.96	
SII	✓	✓			
True Colors					
CSF	✓	✓	✓		✓

The MBTI, had good reliability estimates for four different ethnicities. On the SII, the information is vague. The only specific ethnicities stated were African Americans and Mexican Americans, and both are reported to have "adequately reliable" reliability estimates (Betz and Gwilliam, 2002, as cited in Flores et al., 2006, p. 186). Other research has been done, that concludes that there were greater differences within than between racial and ethnic groups (Fouad, 2002, as cited in Flores et al., 2006, p. 186). The True Colors assessment has no found data on reliability estimates for different races/ ethnicities. The CSF had varying ethnicities tested, but they seemed to group all the scores into one group, "minority ethnicities". Individual ethnicity scores are unknown. However, Schreiner says that racial differences on the CSF were slight (2006, p.8).

In sum, each career assessment test research is found to have various types of reliability and validity. Career counselors have a difficult time to keep on top all the information to better serve their clients. They have to know the appropriate uses of each test, and what age group and ethnicities/races each test is reliable and valid for. It would also be helpful if career counselors understood what aspects of each test are more or less reliable and valid and the implications of

each in a counseling session. Some recommendations for future research are to replicate True Colors research, to conduct longitudinal studies on all the career assessments, and to include the unrepresented ethnicities and races in research.

References

Asplund, M. A., Lopez, S. J., Hodges, T., & Harter, J. (2007). *The Clifton StrengthsFinder 2.0 technical report: development and validation.* Omaha, Neb.: The Gallup Organization.

Betz, N. E., & Gwilliam, L. R. (2002). The utility of measures of self-efficacy for the Holland themes in African American and European American college students. *Journal of Career Assessment,* 10, 283-300.

Buckingham, M., & Clifton, D.O. (2001). *Now, discover your strengths.* New York: The Free Press.

Capraro, R. M., & Capraro, M. M. (2002). Myers-Briggs type indicator score reliability across: studies a meta-analytic reliability generalization study. *Educational and Psychological measurement, 62*, 590-602.

Flores, L. Y., Spanierman, L. B., Armstrong, P. I., & Valez A. D. (2006). Validity of the Strong Interest Inventory and Skills Confidence Inventory with Mexican American high school students. *Journal of Career Assessment,* 14, 183-202.

Fouad, N. A. (2002). Cross-cultural differences in vocational interests: Between groups differences on the Strong Interest Inventory. *Journal of Counseling Psychology,* 49, 283-289.

Harmon, L. W., Hansen, J. C., Borgan, F. H., & Hammer, A. C. (Eds). (1994). *Strong Interest Inventory applications and technical guide.* Palo Alto, CA: Consulting Psychologists Press.

Hodges, T. D. & Harter, J. K. (2005). A review of the theory and research underlying the Strengths Quest program for students. *Educational Horizons, 83*(3), 190-201.

Honaker, S. (2003). *New Implications from convergent validity research with the Myers-Briggs*

Type Indicator. Retrieved August 5, 2009, from True Colors Career Website: http://www.truecolors.org

Kahn, J. H., Nauta, R., Gailbreath, D., Tipps, J., & Chartrand, J.M., (2002). The utility of career and personality assessment in predicting academic progress. *Journal of Career Assessment, 10*, 3-23.

Lopez, S. J., Hodges, T. D., & Harter, J. K. (2004). *Clifton StrengthsFinder technical report: development and validation.* Omaha, Neb.: The Gallup Organization.

Myers, D. G. (2008). *Exploring Psychology, Seventh Edition.* New York: Worth Publishers.

Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *MBTI Manual: A guide to the development and use of the Myers-Briggs Type Indicator* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

Parsons, F. (1909). *Choosing a Vocation* [Electronic version]. Boston: Houghton Mifflin.

Pulver, C. A., Kelly, K. R., (2008). Incremental validity of the Myers-Briggs type indicator in predicting academic major selection of undecided university students. *Journal of Career Assessment, 16*, 441-455.

Reliability and validity of the Myers-Briggs Type Indicator. (n.d.). Retrieved July 10, 2009, from the Myers & Briggs Foundation: http://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/reliability-and-validity.asp

Robinson, C. H., & Betz, N. E., (2004). Test-retest reliability and concurrent validity of the expanded skills confidence inventory. *Journal of Career Assessment,* 12, 407-422.

Schreiner, L. (2006). *A technical report on the Clifton StrengthsFinder with college students.* Retrieved August, 13, 2009, from: https://www.strengthsquest.com/content/111421/Technical-Report-Clifton-

StrengthsFinder-College-Students.aspx

True colors theory. (n.d.). Retrieved August 5, 2009, from: http://www.true-colors-test.com/true-colors-theory.htm

Whichard, J. A. (2006). *Reliability and validity of True Colors.* Retrieved August 5, 2009, from True Colors Career Website: http://www.true-colors.com/TCSite/Reliability%20and%20Validity%20study.pdf

Wilson, D. B. (n.d.). *Interpreting effect size results.* Retrieved August 11, 2009, from: http://www.wilderdom.com/research/meta-analysis.html

Appendix A

Descriptions of Test Themes

Myers-Briggs Type Indicator

Extraverts: A preference to direct energy toward the outer world of people and objects.

Introverts: A preference to direct energy toward the inner world of experiences and ideas.

Sensing: Have a preference of perceiving information through the five senses.

Intuition: A preference of perceiving patterns and interrelationships among information.

Thinking: A preference of drawing conclusions from perceived information by using objective and logical analysis.

Feeling: A preference of drawing conclusions from perceived information by using personal and social values.

Judging: A preference of attitudes toward dealing with the outside world with decisiveness and closure.

Perceiving: A preference of attitudes toward dealing with the outside world with flexibility and spontaneity (Kahn, Nauta, Gailbreath, Tipps, & Chartrand, 2002, p.8).

Stong's Interest Inventory

Realistic: These students like jobs that may focus on mechanical, construction, or repair activities. They like action more than thinking.

Investigative: These students enjoy gathering information, and they like analyzing and interpreting data. They often enjoy science. Working alone is a better option for these students than team efforts, and they do well pursuing advanced degrees.

Artistic: These students need to express themselves creatively. They are usually good with words, and they do well in academic or intellectual environments. They may enjoy arts, music, dramas, or writing.

Social: These students like to work with others. They enjoy sharing responsibilities, caring for others, and teaching, helping, and being the center of attention. They would do well in jobs where they lead, direct or persuade others.

Enterprising: These students talk easily, and are good at selling and leading. They enjoy working with others, and often seek positions of leadership and power. They enjoy taking risks and competing toward goals.

Conventional: These students like paying attention to detail. They often enjoy math and will do well in accounting and investment management. They work well in large organizations, and they do their job with accuracy and organization.

True Colors

Green: The Green personality type can be described as valuing intellectual capabilities. Comfort in these areas creates a sense of personal security and self-esteem for those with Green personalities (True Colors Inc., 2009). Additional characteristics of Green personalities include the need to seek and increase the certainty of personal values through assertiveness. Greens not only express the grounding of theory and data through practicality, logic, and reason, but they constantly seek to understand and solve every problem they encounter (Kalil, 1998).

Blue: The Blue personality type values balance and harmony. Such individuals prefer lives free from tension and are characterized by empathy, openness to aesthetic experiences, and reflective awareness (True Colors Inc., 2009). Blue personalities also place high worth in relationships and feelings of loyalty and belongingness.

Orange: The Orange personality type represents energy, power, and strength. These individuals

feel the will to achieve results, win, and be successful. They desire all things that offer intense and full life experiences. In addition, such personalities generate impulses towards high levels of activity, competition, and productivity (True Colors Inc., 2009).

Gold: Gold personality types can be characterized by responsibility, organization, structure, and the need to fulfill duties. Those with such a personality type value being practical and sensible in their work. They also believe in efficiency, dependability, and embrace the concepts of home and family with fierce loyalty and faithfulness (True Colors Inc., 2009).

Clifton StrengthsFinder

Achiever: These people have a great deal of stamina and work hard. They take great satisfaction from being busy and productive.

Activator: these people can make things happen by turning thoughts into action. They are often impatient.

Adaptability: These people prefer to "go with the flow." They tend to be "now" people who take things as they come and discover the future one day at a time.

Analytical: these people search for reasons and causes. They have the ability to think about all the factors that might affect a situation.

Arranger: These people can organize, but they also have a flexibility that complements this ability. They like to figure out how all of the pieces and resources can be arranged for maximum productivity.

Belief: these people have certain core values that are unchanging. Out of these values emerges a defined purpose for their life.

Command: These people have presence. They can take control of a situation and make decisions.

Communication: These people generally find it easy to put their thoughts into words. They are good conversationalists and presenters.

Competition: These people measure their progress against the performance of others. They strive to win first place and revel in contests.

Connectedness: These people have faith in the links between all things. They believe there are few coincidences and that almost every event has a reason.

Consistency: These people are keenly aware of the need to treat people the same. They try to treat everyone in the world with consistency by setting up clear rules and adhering to them.

Context: These people enjoy thinking about the past. They understand the present by researching its history.

Deliberative: These people are best described by the serious care they take in making decisions or choices. They anticipate the obstacles.

Developer: These people recognize and cultivate the potential in others. They spot the signs of each small improvement and derive satisfaction from these improvements.

Discipline: These people enjoy routine and structure. Their world is best described by the order they create.

Empathy: These people can sense the feelings of other people by imagining themselves in others' lives or others' situations.

Focus: These people can take a direction, follow through, and make the corrections necessary to stay on track. They prioritize, then act.

Futuristic: These people are inspired by the future and what could be. They inspire others with their visions of the future.

Harmony: These people look for consensus. They do not enjoy conflict; rather, they seek areas of agreement.

Ideation: These people are fascinated by ideas. They are able to find connections between seemingly disparate phenomena.

Includer: These people are accepting of others. They show awareness of those who feel left out, and make an effort to include them.

Individualization: These people are intrigued with the unique qualities of each person. They have a gift for figuring out how people who are different can work together productively.

Input: These people have the craving to know more. Often they like to collect and archive all kinds of information.

Intellection: These people are characterized by their intellectual activity. They are introspective and appreciate intellectual discussions.

Learner: These people have a great desire to learn and want to continuously improve. In particular, the process of learning, rather than the outcome, excites them.

Maximizer: These people focus on strengths as a way to stimulate personal and group excellence. They seek to transform something strong into something superb.

Positivity: These people have an enthusiasm that is contagious. They are upbeat and can get others excited about what they are going to do.

Relator: These people enjoy close relationships with others. They find satisfaction in working hard with friends to achieve a goal.

Responsibility: These people take psychological ownership of what they say they will do. They are committed to stable values such as honesty and loyalty.

Restorative: These people are adept at dealing with problems. They are good at figuring out what is wrong and resolving it.

Self-Assurance: These people feel confident in their ability to manage their own lives. They possess an inner compass that gives them confidence that their decisions are right.

Significance: These people want to be very important in the eyes of others. They are independent and want to be recognized.

Strategic: These people create alternative ways to proceed. Faced with any given scenario, they can quickly spot the relevant patterns and issues.

Woo: These people love the challenge of meeting new people and winning them over. They derive satisfaction from breaking the ice and making a connection with another person. (Asplund et al., 2007,pps. 32-36).