1-1-2005

Interpretative implications of revision: calibration of the 1994 and 2004 Strong Interest Inventories

Donna Carla Bailey
Iowa State University

Interpretative implications of revision:

Calibration of the 1994 and 2004 Strong Interest Inventories

by

Donna Carla Bailey

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Psychology

Program of Study Committee:
Lisa M. Larson (Major Professor)
Frederick H. Borgen
Frederick O. Lorenz

Iowa State University

Ames, Iowa

2005

Graduate College

Iowa State University

This is to certify that the master's thesis of

Donna Carla Bailey

has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy

"Perhaps the most valuable result of all education is the ability to make yourself do the thing you have to do, when it ought to be done, whether you like it or not; it is the first lesson that ought to be learned; and however early a man's training begins, it is probably the last lesson that he learns thoroughly."

Thomas H. Huxley (1825 - 1895)

To the members of my thesis committee for helping me learn what, when, and how to do things.

To my beloved Bailey Bunch for helping me learn why they must be done.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

Modern socialization frequently revolves around vocational interests. Often, some of our earliest recollected cognitions center around what we wanted to be when we grew-up. Who among us is unable to recall that we wanted to be a fire chief, or a ballerina, or even the president of the United States of America? As we grow older and enter adolescence, our vocational interests became our foremost concern as the question changes to what we want to do with our lives. Even as adults, one of the most often asked questions upon meeting someone new concerns what vocations we have chosen. Because of the importance placed on vocational interests, it is no wonder that vocational psychology places such substance in interest assessment and career counseling, and that such a variety of tools have been developed to aid counselors in organizing and interpreting client interest and career development.

With its first publication in 1927 (then the Strong Vocational Interest Blank), the Strong Interest Inventory enjoys the longest history of all such tools developed by researchers (Campbell, 1971). Since then, research on the Strong has continuously impacted the areas of interest assessment and career counseling (Borgen, 1986; Campbell & Borgen, 1999; Prince & Heiser, 2000; Strong, 1943; Zytowski & Warman, 1982). Over the past half a century, hundreds of research studies have used the Strong to study changes within an individual, to study characteristic interests of people within specific occupations, to study change in groups and institutions, and to study cognitive processes related to interests, just to name a few (Harmon, Hansen, Borgen, & Hammer, 1994). One of the primary reasons that the Strong enjoys such diversity of domain utility is the strength of both its empirical and theoretical foundations.

The empirical foundation of the Strong is a direct result of the unparalleled empiricism of E. K. Strong Jr. himself (Campbell & Borgen, 1999). Over the course of two decades, Strong's labor culminated in the creation of several categories of clustered occupational scales and a 746-page book containing "197 tables and 52 figures consisting of over 20,000 numbers" (Hansen, 1990). Strong based the derivation of these clustered occupational scales on his belief that interests are liked activities that have a tendency to cluster without regard of specific occupation or industry, and so to form constellations of interests (Strong, 1943). Strong's work and his resulting occupational scales launched the field of interest measurement as we know it today (Campbell & Borgen, 1999), but to maintain its utility in a changing world, the Strong has undergone frequent revision during its almost 80-year history.

Purpose of Test Revision

Currently, no consensus exists for determining the five "W's" of test revision: a) *who* is ultimately responsible for undertaking the revision and who will accept and use it, b) *what* type of revision is needed c) *when* is revision necessary, d) *where* are the areas of the current test in need of revising, and d) *why* is revision even required. The *who, what*, and *where* of test revision varies as a function of the individual measurement tool, and so they must be decided upon by its developers and distributors. However, guidelines for the *when* of test revision are provided by the American Psychological Association (APA) in Standard 3.18 of the *Standards for Educational and Psychological Testing* (APA et al., 1985):

> "A test should be amended or revised when new research data, significant changes in the domain represented, or new conditions of test use and interpretation make the test inappropriate for its intended uses. An apparently old test that remains useful need not

be withdrawn or revised simply because of the passage of time. But it is the responsibility of test developers and test publishers to monitor changing conditions and to amend, revise or withdraw the test as indicated." (p. 29)

The final "W", *why,* is multifaceted. Several authors have attempted to address the complex reasons behind test revision. Reise, Waller, and Comrey (2000) posit revision is a phenomenon of psychometric purity. That is, test revision should often be undertaken to address issues of internal consistency reliability, ability of factor structures to generalize across groups, and construct representation of a test's factor structure (Reise et al.). Adams (2000) takes us from the psychometric to the pragmatic, offering the following reasons for test revision: outdated test materials or stimuli, outmoded language, change in item performance characteristics, shift of the demographic characteristics of the standardization sample, and changes in the definition of problems the test addresses. Perhaps Straus, Spreen, and Hunter (2000) offer the most succinct purpose for revision by stating that a revised test should surpass its predecessor in answering questions about an individuals' functioning. Unfortunately, the path to this seemingly simplistic dogma is wrought with complication.

Challenges of Test Revision

Campbell (1972) himself was one of the first to lament on the challenges of test revision in his cautionary tale to those who were to undertake the revision of the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1940). Campbell's knowledge was hard won by spending over a decade revising the 1966 men's version and 1969 women's version of the Strong Vocational Interest Blank (SVIB; Campbell, 1966, 1969). Specifically, Campbell outlined three obstacles inherent in any revision attempt: a)

technical ignorance, b) user acceptance, and c) logistics. Each of these obstacles will be discussed in turn.

Campbell's first obstacle of test revision, technical ignorance, refers to the difficulty involved in gathering enough data to answer the *what* and the *why* of the five "W's". Accumulating an adequate amount of empirical data to inform decisions regarding what test concepts are no longer useful, what scales are outdated, what language is outmoded, what profiles to use, and what norming groups to focus on (just to name a few) represents a Herculean task. However, before this data gathering task can even begin, *why* a test requires revision must first be clearly and precisely delineated.

The reasoning behind any test revision informs the revision process itself. Campbell (1972) provides four possible reasons a test might require revision: a) items that are out-of-date or unacceptable, b) scales that are out-of-date or no longer descriptive, c) measured traits or variables that are outmoded, and d) a whole empirical keying system that requires replacement. Campbell suggested that if the basis for the revision falls within the two former reasons, then the current form of a measure could be retained with editing of items and scales based on new criterion groups closely resembling the originals. The major advantage of this type of revision process is that disruption of the current system is kept to a minimum (Campbell). This is the type of revision historically undertaken by the developers of the Strong, which helps explain the measure's longevity. That is, if one of the two latter reasons had formed the basis of the need for revision, the Strong would have in effect become an entirely new interest inventory and would have lost its loyal users.

Campbell (1972) speaks directly of this importance of gaining user acceptance for any revision in his second obstacle, maintaining consistency in the system. He adroitly points

out that any reviser expending much effort on a revision is going to want to make at least some detectable changes, else why revise. However, test revision is a delicate balance between making enough changes to ensure widespread use over the older version, and yet not making so many changes that the new version becomes intolerable to users. Campbell offers two strategies to overcome this second obstacle: 1) to achieve a minimum of disruption revisers should make any changes as "adroitly and quietly as possible" (Campbell, p. 121), and 2) revisers should not fix what is not broken, but rather utilize all factors currently working in the original version.

Campbell's (1972) third obstacle in test revision focuses on its logistics, or the practical arrangements involved in the revision process itself. Specifically, he posits three areas of administrative concern: 1) the designation of administrative structure so it is clear who is to do what, 2) provisioning funds needed for the revision, and 3) divvying up credit and royalties for the new revision.

Traditionally, the first of these concerns has been addressed by the establishment of a committee to represent diverse viewpoints. Campbell warns potential revisers of the MMPI against this practice citing the inability of a committee to make the positive, yet potentially risky, changes required in the revision process. Additionally, he argues that a committee is simply incapable of staying intact long enough to handle "a huge diet of what can only be described as menial, grueling work…to complete any large portion of the work" (Campbell, p. 125). Rather, Campbell suggests that only one person (perhaps two or three at the most) must be given absolute authority to succeed or to fail. His reasons for this bold statement are twofold. First, the reviser must have the freedom to make unpopular changes that might go against the status quo. Second, the ultimate responsibility of success or failure lying at the

feet of one individual provides a powerful safety check on the reviser's authority. This responsibility serves to motivate her or him to work tirelessly to determine all possible impacts of potential changes.

To illustrate the ramifications of even a seemingly trivial decision change on the reviewer, Campbell (1972) recounts the ensuing problems with the choice of the test booklet color for the men's and women's revisions of the SVIB. Campbell and his crew decided to choose a color from amongst those the printer had in stock, gray, brown, pink, and blue. They chose blue for the men's revision as they felt it livelier. Two years later, when it came time to choose a color for the women's revision, they were naturally left with gray, brown, and pink. Realizing it might be "overly cute", they chose pink for the women's version finding it a preferable alternative to gray or brown. Most people took the color changes with good humor, but Campbell lost a valuable data source when an executive secretary for a company that employed 800 female high-level managers (a much needed and rare data source), took exception to the stereotypical color of the test booklets. This example crystallizes the importance of considering the impact of even seemingly trivial changes.

Campbell's (1972) second area of administrative concern is the difficulty inherent in recruiting possible funding sources. The viability of any test revision rests to a large degree on its cost. Once a potential budget is determined, possible funding sources must be identified. Some potential sources include the following: federal funds from government agencies such as the National Institute of Health (NIH) or the National Institute of Mental Health (NIMH) for non-commercially viable tests, commercial scoring services, test publishers, and various organizations that use the measure in some capacity (Campbell).

Also speaking to economic concerns is the problem of remuneration for those working on the revision. This is Campbell's (1972) third area of administrative concern, divvying up the rewards. Though often not discussed overtly, issues of authorship and royalties can become a hotbed of interpersonal strife. Campbell (p. 129) offers no concrete solution to this concern reporting, "On the Strong, we just blundered along until everyone felt at least comfortable enough not to speak out. In retrospect, this casual approach was probably as good as any." Though Campbell's advice on the difficulties of revision was directed towards the potential revisers of the MMPI, many of these issues were considered during the next three rounds of the revision of the Strong. The purpose and result of each revision round is discussed next.

The 1960's and 1970's Revisions

Though the original occupational scales proved an empirical feat, they were difficult for the layperson without the requisite special knowledge necessary for correct interpretation to fully understand. The cause of this difficulty lay in their heterogeneous nature; the items do not all correlate highly within a general group of people (Campbell & Borgen, 1999). Thus, a counselor would know that a high score on a scale is indicative of a person sharing the relative likes and dislikes of others within an occupation, but the content of the scale itself was not always so apparent. This difficulty in interpretation indicated a need for revision to increase the Strong's user-friendly capability (Campbell & Hansen, 1981).

The first major conceptual change in the Strong occurred in the 1960's with the introduction of content scales meant to enhance the breadth of utility of a measure known previously only for its occupational scales. This was Campbell and his associates' addition of the 23 Basic Interest Scales (Campbell, Borgen, Eastes, Johansson, & Peterson, 1968). The

introduction of these homogeneous content dimensions greatly enhanced the meaning and interpretability of the Strong (Campbell & Borgen, 1999), and resulted in the publication of the Men's Form (SVIB-M) in 1966 (Campbell, 1966) and the Women's Form in 1969 (Campbell, 1969). It was during this revision that Campbell garnered the experience and growing pains which allowed him to expound upon the trial and tribulations of test revision in his 1972 warning to the future revisers of the MMPI (see above).

An equally momentous revolution in the Strong was the addition in the 1970's of Holland's theoretical system. The increasing complexity of the measure, due in part to the addition of the Basic Interest Scales, helped pave the way for the inclusion of another major set of content scales, the new "Holland Scales", resulting in a paradigm shift in future applications of the Strong (Campbell & Borgen, 1999). Building on Strong's ideas and multivariate insights, Holland developed a useful, practical occupational classification system of six occupational themes (Holland, 1966). Soon after this, Campbell noticed patterns appearing in his Occupational and Basic Scales that approximated Holland's hexagon (Campbell & Borgen, 1999). In 1972, in a striking blending of paradigms, Campbell and Holland demonstrated that Holland's theory provided a useful framework for organizing Strong's data, and they presented the newly created Holland scales for the SVIB-M (Campbell & Holland, 1972). Hansen and Johansson (1972) presented the newly created Holland scales for the SVIB-W.

The last major revision of the Strong during these decades took place in 1974 and resulted in a name-change from the SVIB to the SCII, or to the Strong-Campbell Interest Inventory (Campbell, 1974). This revision reflects all four of Campbell's major innovations to the Strong: 1) the inclusion of the Basic Interest Scales; 2) the introduction of the Holland

scales; 3) the reorganization of the Strong around the Holland hexagonal scheme; and 4) the merger of the men's and women's inventories into a single measure (Borgen, 1986). It was also in this 1974 version that the General Occupational Themes (GOT's) based on the six new Holland RIASEC (Realistic, Investigative, Artistic, Social, Enterprising, and Conventional) typology scales resulted in an underlying theoretical structure for the Strong (Campbell & Holland, 1972). Throughout each of these revisions, careful attention was paid to follow Campbell's sage advice. That is, disruptive changes due to the addition of the Basic Interest Scales and the inclusion of Holland's RIASEC as a conceptual framework were minimized whenever possible, and the highly effective and useful occupational scales were retained. Refer to Table 1.1 below for a description of the RIASEC themes.

Table 1.1

Brief Descriptors Associated with Each of the Six General Occupational Themes (RIASEC)

THEME	INTERESTS	WORK ACTIVITIES
Realistic	Machines, tools, outdoors	Operating equipment, using tools, repairing
Investigative	Science, theories, ideas, data	Performing lab work, solving abstract problems, researching
Artistic	Self-expression, art appreciation	Composing music, writing, creating visual art
Social	People, team work, human welfare, community service	Teaching, explaining, helping
Enterprising	Business, politics, leadership, influence	Selling, managing, persuading
Conventional	Organization, data, finance	Organizing, operating computers

Note. Adapted from *Strong Interest Inventory: Applications and Technical Guide*. Harmon, Hansen, Borgen, & Hammer (1994).

The 1981 and 1985 Revisions

In 1981, Campbell and Hansen revised the Strong in an effort to continue to improve the measure's empirical foundation. This revision also attempted to increase the breadth of possible occupations that could be considered viable for both sexes. This presented an attempt to update previous versions which included a wide range of occupations considered viable for only one sex or the other (Hansen, 1986).

Two significant changes mark the 1985 revision of the Strong. The first of these occurred through the inclusion of new nonprofessional occupational scales in an attempt to respond to complaints that the utility of the Strong lay only in assessing college-educated individuals (Hansen, 1985). The cause of this criticism stemmed from the hitherto lack of inclusion of any vocational-technical and nonprofessional occupations represented in the occupational scales. The inclusion of these new nonprofessional scales allowed the Strong to have interpretive implications for a greater array of occupations. Table 1.2 below presents the 17 vocational/technical and nonprofessional Occupational Scales added to the 1985 version of the Strong.

Table 1.2

Vocational-Technical and Nonprofessional Occupational Scales Added to the 1985 SII

1985 SII Additional Occupational Scales
1. Air Force Enlisted Personnel
2. Army Enlisted Personnel

Table 1.2 (Continued)

3. Bus Driver
4. Carpenter
5. Chef
6. Electrician
7. Emergency Medial Technician
8. Florist
9. Food Service Worker
10. Funeral Director
11. Horticultural Worker
12. Marine Corps Enlisted Personnel
13. Medical Technician
14. Navy Enlisted Personnel
15. Optician
16. Respiratory Therapist
17. Travel Agent

Note. Adapted from the *Manual for the Strong Interest Inventory*. Hansen & Campbell (1985).

The second significant change involved the development of new general reference samples chosen randomly from a re-normed set of general reference samples of 300 males and 300 females (Hansen, 1985). The re-norming of this General Reference Sample provided increased scaling accuracy both within and across sexes. The main impact of this change was

to decrease sex bias while still maintaining the Strong's theoretical structure. After the 1985 revision, the Strong was unchanged until 1994.

The 1994 Revision

The 1994 317-item Strong represented an attempt to modernize its theoretical scales (GOTs and BISs) by deleting and replacing any outdated scales, and to revise and tighten the psychometric properties of any remaining scales (Harmon et al., 1994). The researchers were guided by an item-based factor analysis that revealed that the majority of BIS factors were in reality sub-factors of the GOTs and that 75% of the items found on the BISs compose the GOTs. These findings led to the extension of the Basic Interest Scales from 23 to 25, the creation of four new scales (Applied Arts, Culinary Arts, Computer Activities, and Data Management), and increased the interpretation of the individual as being dominant in one or more GOTs while having additional BIS descriptors (Harmon et al., 1994).

Beyond modernizing and revising the GOT and BIS content scales, the 1994 revision included four new Personal Style Scales (PSS): Work Style, Learning Environment, Leadership Style, and Risk Taking/Adventure. Similar to the GOTs and BISs, the PSSs contain standard scores based on the General Reference Sample of the 18, 951 individuals tested for this revision (Harmon et al., 1994). However, these scales do differ from the GOTs and BISs in that they are constructed as bipolar scales with differing interpretations attached to either pole (i.e., a score on the lower pole of the Risk Taking/Adventure scale would infer a preference to play it safe; while a score on the upper pole of this scale would indicate a preference for taking chances).

The researchers of the 1994 Strong revision were guided by a factor analysis, and performed Cronbach alpha reliability analyses in an effort to increase the reliability of each

homogeneous scale (Harmon et al., 1994). The 1994 Strong demonstrated good reliability and validity. Further information regarding the 1994 Strong is presented in the Method's section under Instruments.

The 2004 Revision

This latest revision of the Strong, with 291-items, represents a substantial update from the 1994 version, and it embodies the researchers' attempts to capture the evolving occupations and work activities of the 21st century. A new norming sample, now called the General Representative Sample (GRS) rather than the General Reference Sample was taken. Specifics regarding this new GRS are presented in Chapter Two (the Methods section) under Instruments.

It is important to note that the present study did not use the actual 2004 Strong booklets, as they were not yet released during the periods of active data collection. Rather, this study made use of the research version of the inventory which contained 361-items, including 70 new items designed to measure contemporary interests that were included in the final 291-item version (i.e., computer hardware and programming). Specific item edits, additions, and deletions between the 1994 and 2004 SIIs can be found in the 2004 Strong manual (Donnay, Morris, Schaubhut, & Thompson, 2005).

Table 1.3 below presents a global overview of the structural and item changes of the 2004 SII. The 2004 Strong contains three types of content scales that will be included in this study: six General Occupational Themes; 30 Basic Interest Scales; and five Personal Style Scales (representing the addition of the Team Orientation PSS). The Occupational Scales were also retained in this latest revision of the SII, but will not be specifically discussed as they are not included in the present study.

Table 1.3

Global Structural and Item Changes between the 1994, 2004 Research Version, and 2004 Strong Interest Inventories

1994 Strong		2004 Strong (Research and Final Versions)	Research Version		Final Version	
Parts	**Total Items**	**Sections**	**Total Items**	**New Items**	**Total Items**	**New Items**
Part 1: Occupations	135	Section 1: Occupations	124	13	107	11
Part II: School Subjects	39	Section 2: Subject Areas	50	17	46	16
Part III: Activities	46	Section 3: Activities	119	83	85	59
Part IV: Leisure Activities	29	Section 4: Leisure Activities	34	14	28	10
Part V: Types of People	20	Section 5: People	20	1	16	0
Part VI: Preference Between two Activities*	30	Section Not Applicable	0	0	0	0
Part VII: Your Characteristics	12	Section VI: Your Characteristics	14	3	9	2
Part VIII: Preference in the World of Work	6	Section Not Applicable	0	0	0	0
Totals	**317**		**361**	**131**	**291**	**98****

Note. Adapted from Strong Interest Inventory Manual: Research, Development, and

Strategies for Interpretation. Donnay et al. (2005). * *Denotes 15 items modified into Likert-type items for the 2004 Strong.* ** *Denotes that 20 of the 98 new items on the 2004 Strong represent only minor edits in wording.*

Perhaps the most dramatic change to the 2004 version of the Strong is the introduction of the 5-point response format to replace the 3-point LID format of the 1994 SII. Subjects now have the option of responding on a 5-point Likert scale ranging from "Strongly Like" to "Strongly Dislike" rather than the "Like", "Indifferent", and "Dislike" response choices on all previous versions of the Strong. Once again, the researchers wrote new items in new content domains and used a factor analysis to guide the identification of specific dimensions of interest, which allowed the expansion of the domains of interest in the new 2004 Strong (Schaubhut, Donnay, Gasser, & Borgen, 2004). Examples of new content dimensions in the revision include computer technology and entrepreneurship. Further information regarding the new 2004 Strong is provided in Chapter Two.

Reliability

The psychometric stability of the Strong through all of its revisions is aided by the robustness of vocational interests (Borgen, in press). For example, college students who enjoy mathematics were most likely middle school students who enjoyed mathematics. Thus, the content scales of the Strong (GOT, BIS, PSS) all have high internal consistency, with the six Holland RIASEC scales in the 1994 version all demonstrating alpha reliabilities of .90 or higher (Harmon et al., 1994).

The reliability of the 1994 Strong is well established. Harmon et al. (1994) used four samples, taken over an interval of one to three months, to evaluate the test-retest reliability of the 1994 Strong. The four samples (A through D) consisted of one sample taken from a

population of working adults, and the remaining three samples taken from a population of university students at three different universities. The test-retest reliabilities for these samples are reported in summary format in chapter 6 of the Strong *Interest Inventory Applications and Technical Guide* (Harmon et al.), and show high reliabilities for all four samples. The median correlations (and their ranges) for samples A through D respectively are as follows: .90 (.80 to .95), .87 (.70 to .93), .85 (.71 to .96), and .84 (.66 to .96). Swanson and Hansen (1988) further established the test-retest reliability of the 1994 Strong by correlating current scores to archival scores collected 12 years previously. Their results indicated high reliability showing a correlation of .72 for the 242 females tested, and a correlation of .73 for the 167 males tested.

Research centering on the homogeneity of the scales on the 1994 Strong have supported its high internal consistency (Betz & Taylor, 1982; Hansen & Campbell, 1985; Tracey & Rounds, 1993). As reported by Harmon et al. (1994), Cronbach alphas for the six GOTs in respective RIASEC order were as follows: (.93, .91, .94, .90, .90, .90).

The high reliability shown for the 1994 Strong translates into user loyalty. Thus, it is imperative that equal or improved reliability is demonstrated for the revised 2004 Strong to address Campbell's (1972) second obstacle of test revision, user acceptance. The process of attempting to demonstrate equal reliability between parallel measures in often referred to as test calibration. The main purpose of this study is to calibrate the 1994 and 2004 Strongs in an effort to aid in the acceptance of the 2004 Strong. However, several authors have elucidated difficulties intrinsic to the calibration process.

Smith and McCarthy (1995) posited five fundamental psychometric criteria which should be considered along all stages of the revision process. Thus, during the process of

calibrating a newly revised measure with prior versions the following five criteria should be attended: 1) the hierarchical structure, 2) internal consistency reliability, 3) the content homogeneity of underlying facets to ensure equal construct representation, 4) item discriminatory power, and 5) replication of factor structure (Smith and McCarthy).

Strauss, Spreen, and Hunter (2000) outlined another problem that can plague the test revision process. This problem occurs as a result of the addition of a substantial number of new items and/or new subtests. Too many new items or subtests can result in a lack of applicability of decision rules based on the older version to the newer revision. The authors cite the revision of the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981) resulting in the Wechsler Adult Intelligence Scale-III (WAIS-III; Tulsky, Zhu, & Ledbetter, 1997) as an example of this first problem. The more than 40% of new items in each subtest resulted in difficulty in comparison of scores between the two versions (Strauss et al.). The 2004 Strong retained two-thirds of the items from its 1994 predecessor resulting in 98 of its 291 items being new. However, 20 of the 98 new items on the 2004 Strong represent only very minor changes in wording (e.g., from the 1994 item number 79 "Manager, child care center" to the 2004 item number 65 "Manager of child care center"). Therefore, only 26.8% of items on the 2004 Strong represent substantial changes from the 1994 version. However, despite this much reduced number of relatively new items, the impact of the new items on the current interpretive framework must be established. Thus, a second purpose of the present study is to determine the interpretive implications of the 2004 revision.

Validity

Though the establishment of the reliability of the Strong is important, demonstrating its validity is paramount to ensuring that the information provided by the measure accurately captures the vocational interest construct. A reliable measure may indeed produce replications of scores, but it is its validity that renders the scores meaningful. The validation of the Strong is a necessary component of incorporating its results into career decision-making, and the numerous studies examining the validity of the Strong speak to its usefulness as a vocational assessment tool.

Early Validity Research. The earliest study on the predictive validity of the Strong (then the SVIB) was conducted by E. K. Strong himself (1943). Strong performed a ten-year test-retest study on both a sample of 140 male college seniors and a sample of male college freshman. For both samples, Strong found greater-than-chance scores, though interest prediction for the younger sample showed slightly less accuracy. Based on these samples, Strong concluded that men who continued in a given occupation obtained higher scores for their chosen occupation relevant to other occupations, scored higher than men in occupations other than their own, and scored higher than men who had changed occupations during the ten-year period (1943).

Strong (1955) conducted a follow-up study which further demonstrated his measure's predictive validity. In this study, Strong sampled 600 male students at Stanford University who met the selection criteria of having an occupation at the time of follow-up that matched with an Occupational Scale on the Strong. The straightforward methodology of this study was to compare the men's current occupations to their college scores yielded 18 years previously. In Strong's conclusions, he reported expectancy ratios of 3.6 chances to 1 that a man achieving an "A" rating (or achieving a score of 45 or higher on the relevant OS) on a

specific occupation would enter that occupation, and a ratio of 5.0 to 1 that a man with a "C" rating (or achieving a score of 25 or lower on the relevant OS) would not enter that occupation (1955).

Validity Research on the 1960's and 1970's Versions. Dolliver (1975) was one of the first researchers to examine the concurrent validity of the GOTs, BISs, and OSs on the men's version of the SVIB (Campbell, 1966). Dolliver examined validity hit rates by defining a hit as a score at or above 55 for the GOTs, at or above 58 on the BISs, and at or above 45 on the OS with the most relevancy to the participant's current occupation. Dolliver reported hit rates of 39% for the GOTs, 41% for the BISs, and 30% for the OSs for predicting occupational membership after adjusting for chance.

Cairo (1979) used Dolliver's (1975) criteria to further demonstrate the concurrent validity of the GOTs and BISs of the SVIB, but diverged from Dolliver's study by collecting additional information from each of the 36-year old men participating in their longitudinal study of career development. The increased specificity allowed by the additional information concerning the participants' occupational titles and job descriptors prompted Cairo to conclude that this increased precision of data accounted for his higher hit rates of 47% for the GOTs and 44% for the BISs in predicting occupational membership.

Worthington and Dolliver (1977) examined the validity of both the SVIB and the SCII (Campbell, 1974). Rather than using Dolliver's previous method (1975) of assessing hit rates for the content scales in predicting occupational membership, the authors used McArthur's (1954) method of identifying the OS with the closest match to the participant's current job, and then making a judgment concerning whether the OS was a direct or indirect measure of the participant's interest in the current occupation. Based upon this criterion, the

authors found predictive validity hit rates for the SVIB to be 42%, and concurrent validity hit rates of 29% and 39% percent respectively for the SVIB and SCII after accounting for base rates.

The validity of the SCII was also the focus of investigation by Spokane (1979) who examined the concurrent and predictive validity of its GOTs with a sample of female (n = 157) and male (n = 304) college students completing the measure both before entering their first year and again during their senior year of college. Spokane identified a predictive validity hit as an identical match of a senior's occupational preference with the highest GOT she or he identified four years previously, finding rates of 34.4% for women and 39.7% for men. Concurrent validity hit rates were defined as an identical match between a senior's occupational preference and her or his highest senior GOT code, and were reported as 34.4% for women and 43.6% for men. Spokane reported only partial support for the predictive validity of the SVIB's GOTs stating modest differences between women and men.

Employing a new approach for examining validity, Borgen (1972) used univariate and discriminant methods to test the predictive ability of the OSs and BISs of the SVIB. Participants in the study included 780 male National Merit Scholars who were subsequently separated into a validation sample (n = 511) and a cross-validation sample (n = 269). Borgen concluded only minor differences in predictive ability between the two sets of scales finding a direct hit rate of 23% for the OSs and 24.5% for the BISs in predicting occupational choice in the cross-validation sample. The author again noticed this minor difference after dichotomizing the career groups into science versus non-science fields reporting hit rates of 69.1% and 72.5% for the OSs and BISs, respectively.

Validity Research on the 1980's Versions. The 1981 and 1985 versions of the SII each represent a revision of the 1974 Strong-Campbell Interest Inventory (SCII). As such, much of the validity studies cited in the 1985 *Manual for the Strong Interest Inventory* (Hansen & Campbell, 1985) refer to research conducted on the 1974 predecessor rather than on the 1981 or 1985 revisions. According to Hansen and Campbell, because the SCII scales are based directly on the earlier scales of the Strong Vocational Interest Blank (SVIB), the data for the earlier version retains it relevancy.

For example, Varca and Shaffer (1982) conducted a study using both adolescent and adult participants. Their findings showed that the GOTs were successful in predicting the avocational and leisure activities participants choose to engage in, and additionally demonstrated that avocational interests, like their vocational interest counterparts, are stable over time. In a similar fashion, Johnson and Johnson (1972) using a college sample, showed that consistency in scores on the BISs and OSs led to a greater degree of predictive accuracy. That is, students who scored high on both the sales-area Occupational Scale and the Sales Basic Interest Scale were more likely to enter a vocation within the sales domain than students scoring high on only one of these scales.

Validity Research on the 1994 Version. Many researchers have also demonstrated the validity of the 1994 Strong. The validity of all three types of content scales (GOTs, BISs, and PSSs) has been evaluated with positive results. Initial validity studies by Harmon et al. (1994) demonstrated the concurrent and construct validity of the GOTs by examining the rank order means of each of the occupations within a GOT. For each GOT, the occupations aligned in a theoretically expected way. Harmon et al. also examined the GOT's ability to

differentiate between college majors, and again found that the college majors grouped according to Holland's typology as expected.

The concurrent validity of the Strong has also been evaluated against similar interest measures. Both Donnay and Borgen (1996) and Olsen (1996) successfully demonstrated its concurrent validity by using multivariate methods to predict occupation from the three content scales, and by showing the similarity of these scales across sex, respectively. Specifically, Donnay and Borgen (1996), using combined-sex samples of the 50 occupations in the General Reference Sample, found direct hit rates of 10.4% for the GOTs (five times greater than chance), hit rates of 21.8% for the BISs (ten times greater than chance), and hit rates of 8.6% for the PSSs (four times greater than chance).

Olsen (1996) examined the concurrent and predictive validity of the content scales (GOTs, BISs, PSSs) of the 1994 SII for the same 17,165 participants selected from the General Reference Sample as Donnay and Borgen (1996), but separated the sample into three different same-size groups (combined-sex, women, and men) representing 43 occupations. Using univariate techniques, Olsen concluded all 35 scales contributed significantly to the prediction of occupational group separation ($p < .005$) for all three groups, reporting Wilkss's lambdas for ranging from .655 for the Work Style scale to .974 for the Culinary Arts scale. The author used discriminant analyses for the multivariate prediction of occupational group membership concluding that as the number and specificity of variables increased, the predictor sets' ability to differentiate among occupational groups increased reporting Wilks's lambdas ranging from .456 for the combined-sex sample to .062 for the men's sample. Using Cohen's (1988) standard of interpreting a lambda of .86 or less as having a large effect size, the high validity of the Strong is readily apparent.

Donnay and Borgen (1999) examined the concurrent validity of the GOTs with the General Confidence Themes (GCTs) of the Skills Confidence Inventory (SCI; Betz, Borgen, & Harmon, 1996) by measuring the vocational interests and vocational self-efficacy of 1,105 employed women and men. Their results replicated a study by Tracy (1997), and both studies found that the measures offered distinct but similar prediction of satisfied occupational membership. Additionally, Hansen (1986) examined the concurrent validity of the Strong and the Vocational Preference Inventory (Holland, 1985) and reported that the measures appear to delineate occupations according to Holland's theoretical predictions.

Harmon et al. (1994) examined the concurrent validity of the BISs in discriminating among occupational groups. They demonstrate the validity of the BISs by offering graphic representation of the mean scores of participants employed in various occupations. Results showed that occupational group members obtained the highest scores on those occupations most closely resembling their current occupation of choice. For those occupations unrelated to their occupation of choice, participants tended to score at or below average levels. However, perhaps the greatest proof for the concurrent validity of the BISs is the lack of overlap between scales (Harmon et al.).

Savickas, Taber, and Spokane (2002) assessed the convergent and discriminant validity of similarly and like-named scales on five different interest inventories: (a) the Campbell Interest and Skills Survey (CISS; Campbell, Hyne, & Nilsen, 1992)., (b) the Kuder Occupational Interest Survey-Form DD (Kuder & Zytowski, 1991), (c) Holland's Self-Directed Search (Holland, 1994), (d) the Strong Interest Inventory-Skills Confidence Inventory Edition, and (e) the Revised Unisex Edition of the ACT Interest Inventory (UNIACT; Swaney, 1995). For purposes of this study, only the method correlations

corresponding to the SII are reported. The participants in this study were given all five measures, and the group consisted of 80 women and 38 men all of whom were employed as career counselors or professors. The authors used multitrait-multimethod matrices to show moderate correlations of similar and like-named scales to indicate convergent and discriminant validity.

Specifically with regard to convergent validity, Savickas et al. (2002) reported median monotrait-heteromethod correlations of $r = .55$ for the Realistic scale, $r = .76$ for the Investigative scale, $r = .55$ for the Artistic scale, $r = .59$ for the Social scale, $r = .64$ for the Enterprising scale, and $r = .47$ for the Conventional scale. Only two matched scales failed to find significance; the Artistic scale did not significantly converge with the Kuder Occupational Interest Survey-Form DD ($r = .05$, $p > .01$) and the Conventional scale did not significantly converge with the Campbell Interest and Skills Survey ($r = .18$, $p > .01$).

Savickas et al. (2002) also reported heterotrait-heteromethod correlations which were used as a control against the monotrait-heteromethod correlations, to assess convergent validity as described above, and to determine the discriminant validity of the Strong. The authors reported proportionately significantly fewer control values (heterotrait-heteromethod correlations) than validity values (monotrait-heteromethod correlations), with only 33 of the 145 control values showing statistical significance ($z = 7.28$, $p < .001$). Based upon these correlations, the authors concluded the SII showed good discriminant validity.

Schmidt, Lubinski, & Benbow (1998) evaluated the construct validity of the Strong and the Study of Values (SOV; Allport, Vernon, & Lindzey, 1970) for a sample of 695 intellectually gifted 13-year olds by examining both the intercorrelations and the external relation across 59 relevant external criteria. The authors concluded that most of the Strong

and SOV scales covaried as expected, with Religious Activities and Religious values obtaining the highest correlation (r = .70). Additionally, construct validity for gifted youth was also supported by multiple correlations of the SII and SOV to external criterion finding that for all criterion variables, the Strong displayed correlations with the external criteria that were comparable or greater than those with the SOV.

Due to the empirical base of the Strong, criterion-related validity is the largest player in determining whether the measure correctly predicts the behavior that it was designed to predict. For example, does a respondent become a banker if their profile shows that they share interests with other bankers? Research has shown strong support for the measure's ability to predict educational aspirations (Rottinghaus, Lindley, Green, & Borgen, 2002), to predict college major (Issacs, Borgen, Donnay, & Hansen, 1997; Ralston, Borgen, Rottinghaus, & Donnay, in press), and to predict occupational membership (Donnay & Borgen, 1996; Lattimore & Borgen, 1999).

Validity Research on the 2004 Version. Due to its very recent release, few studies exist on the validity of the 2004 Strong. One study by Gasser (2005) was found which looked at the concurrent validity of the 2004 Strong for identifying college students in 31 different majors. The sample used in this study is of special import to the current study as it was used as a comparison sample along with the 2004 GRS to test the representativeness of the current sample. Gasser used a 1,873 college student subset of the participants garnered by CPP, Incorporated as part of the 2004 Strong revision data collection effort. Of the 1,873 participants, 1,403 were women and 470 were men with a mean age of 21.93 (SD = 5.6). The ethnic breakdown of the sample was as follows: 73% Caucasian, 11% African-American, 8% Hispanic, 5% Asian, 3% other or multiple ethnicities (Gasser).

Specifically, Gasser (2005) examined the ability of the 41 content scales of the 2004 Strong to identify college students in 31 different college majors. Multivariate predictive analysis resulted in discriminatory hit rates of 33.7% for the BISs alone, 15.5% for the GOTs alone, 12.9% for the PSSs alone, and a combined hit rate of 38.3%. These findings support previous research on the 1994 Strong showing the BISs to have the greatest predictive ability for college major (Donnay & Borgen, 1996).

The Present Study

As with each new version of the Strong Interest Inventory, there is an accompanying need for research to ascertain the equivalency of the new 2004 version with its predecessor, and in particular to access issues of reliability and validity. This need is clearly specified by the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), "a clear rationale and supporting evidence should be provided for any claim that scores on different forms of a test may be used interchangeably" (Standard 4.10; p. 57). The empirical and theoretical foundations of the Strong both aid in increasing its reliability and validity, and have been well established for the 1994 version (Donnay & Borgen, 1996; Olsen, 1996). However, to gain confidence in the results of the new 2004 revision, it is important to calibrate the ability of the 2004 scales with similar names and content to the 1994 scales to yield equivalent outcomes. A test of the validity of the 2004 Strong is beyond the scope of this study. Rather, this study is an attempt to calibrate each of the parallel content scales of the 1994 and 2004 versions of the Strong to aid in establishing the reliability of the 2004 revision. Additional research is required to establish the validity of

the 2004 Strong. Such research has begun (Borgen, Larson, Bailey, Thompson, & Donnay, 2004; Gasser, 2005).

Preliminary Analyses

The first preliminary analysis will help determine the representativeness of the current sample by comparing the means and standard deviations of the 41 analogous content scales in the current sample to two comparison samples. The 2004 General Reference Sample will serve as the first comparison sample and a recently conducted study on the concurrent validity of the 2004 Strong by Gasser (2005) will serve as the second comparison sample. The purpose of the second preliminary analysis is to provide a test of the normality assumption. The third preliminary analysis will address the fourth of Smith and McCarthy's (1995) criteria by testing the ability of the 1994 and 2004 Strongs to discriminate between sexes, and will test the equivalency in magnitude of observed sex differences between the two versions.

Hypotheses

Specifically, this study examined the following two main hypotheses:

1. Paired sample t-tests will show no meaningful differences between the means of the parallel content scales of the 1994 and 2004 Strong Interest Inventories. Specifically, no mean differences greater than one-half standard deviation are expected.

2. Strong positive relations will be present between the analogous content scales of the 1994 and 2004 Strong Interest Inventories. It is important to determine how well the content scales across the 1994 and 2004 Strongs calibrate in order to ensure that the interpretation of the scales for the 2004 version closely approximate the interpretation of the scales using the 1994 version. For example, persons' responses on the 2004 Realistic GOT

should not differ dramatically from their responses on the 1994 Realistic GOT. Specifically, the analogous content scale Pearson product moment correlations will all be greater than or equal to .85; the lowest correlation found in the 1994 revision.

This second hypothesis also addresses the second purpose of this study, to determine the interpretive implications of the 2004 revision, by showing that individual's responses on the 2004 Strong are highly correlated with their 1994 responses. Note: originally, the author intended to demonstrate the content homogeneity of the parallel content scales of the 1994 and 2004 Strongs (the third criterion) by using Cronbach's alpha's as a measure of internal consistency. However, due to the proprietary nature of the item-level scoring of the Strong, the author was unable to complete this analysis.

Rationale for Hypotheses

The rationale for both hypotheses follows first from a practical issue. Equivalence between parallel forms of measures must be established in order to rely on all the prior research and counseling information garnered from previous versions. There are seven decades of research on the Strong Interest Inventory – can we rely on it? Second, Campbell (1972) himself lamented on some of the important considerations and obstacles of test revision. It is beyond the scope of this study to test his first obstacle (technical ignorance) or his third obstacle (practical administrative arrangements). Therefore, this study attempts to aid in overcoming his second obstacle (user acceptance) by calibrating the 2004 revision with its popular 1994 predecessor. Moreover, the first hypothesis addresses the equivalency of the hierarchical conceptual structure (Smith & McCarthy, 1995) of the two versions.

CHAPTER 2. METHOD

This section begins by describing the sampling population and presenting the procedures followed during the data collection process. Next, the 1994 and 2004 Strongs are each described in detail, with emphasis placed on applicable divergent areas between the two versions. A narrative of each set of content scales (i.e., the GOTs, BISs, and PSSs, respectively) is then presented. This is followed by an outline of the specific hypotheses tested in this study. Finally, the design analyses conducted in this study concludes the section.

Participants

The participants in this study are from data collected from a large Upper Midwestern university during the fall semester of 2002 and the spring semester of 2003. There were 355 participants in the fall 2002 data collection pool and 272 participants in the spring 2003 data collection pool for a total of 627 participants. All participants were volunteer undergraduate students who completed a research version of the combined 2004 Strong containing all 317-items of the 1994 Strong and the research items considered for the 2004 Strong. Participants received course extra credit for their participation, with numerous alternate research options existing for students to gain this course extra credit.

The fall 2002 sample consisted of 129 males and 226 females (four participants did not indicate sex). Of the 355 participants, 90.4% were White, non-Hispanic, 1.4% were Hispanic-American, 2% were Asian-American, 2.8% were African-American, 1% marked "other", and 2.4% did not indicate ethnicity. The mean age of the sample was 19.27 years. The spring 2003 sample consisted of 272 participants with 119 males and 148 females (five participants did not indicate sex). Of the total participants, 81% were White, non-Hispanic,

4.4% were Hispanic-American, 3.6% were Asian-American, 4% were African-American, 3.3% marked "other", and 3.7% did not indicate ethnicity. The mean age of the participants in this sample was 20.09 years. These two data sets were merged to create one data set consisting of 631 participants with 248 males and 374 females (a total of nine participants did not report sex). The total mean age of the merged data set was 19.33 years with a standard deviation of 1.55. The ethnic breakdown of the total data set is as follows: 86% White, non-Hispanic, 2.7% Hispanic-American, 2.9% Asian-American, 3.2% African-American, 2.2% international students, and 2.1% of participants marking other.

Procedures

All participants in this data collection were undergraduate students who volunteered to be contacted for further research opportunities during mass testing sessions. The participants were contacted by email and offered the chance to participate in this study in exchange for course extra credit. Participants wishing to take part arrived at a pre-selected location, were led into a room, and then were given a packet containing an informed consent form, a short demographic questionnaire, and the research version of the combined 1994 and 2004 Strongs.

Students completed a 428-item combined 1994/2004 research packet. The packet included all 317 items of the 1994 Strong, the 78 new items of the 2004 Strong, and an additional 33 new items of the 2004 research version that were ultimately not included in the new 2004 Strong. For a global overview of structural and item changes between the 1994, research version of the 2004, and final version of the 2004 Strongs, please refer to Table 1.3.

A complete listing of all retained, revised, additional, and edited items can be found in the new 2004 Strong manual (Donnay et al., 2005).

Two aspects of the research version of the Strong used in the present study are of special note. First, all items that appeared on both the 1994 and 2004 Strongs were included only once. Second, all items were answered on the new five-point response format of the 2004 Strong. This second ramification translated into a need to collapse the five-point response format to a three point L-I-D response format for the 1994 in order to use its traditional scoring routine. That is, the first two responses were collapsed into "Like", the third response was retained and represented "Indifferent", and responses four and five were collapsed to represent "Dislike".

Additional measures were collected including the Multidimensional Personality Questionnaire (MPQ; Tellegen, 1982) and the Expanded Skills Confidence Inventory (ESCI; Betz, Borgen, Rottinghaus, Paulsen, Halper, & Harmon, 2003), but they were not a part of this study and so are not discussed further. The research version of the combined 1994 and 2004 Strongs, the MPQ, and the ESCI were all counterbalanced in an attempt to control for order effects. The researchers read the informed consent and provided additional instructions for completing the packet. The participants then completed the packets during a three-hour interval, returned them to the researchers, and were given their extra credit and debriefing forms. These procedures were identical for all data collection sessions over both semesters.

In an effort to test the null hypothesis that no statistically significant differences between counterbalanced groups would exist due to participant fatigue, one-way Analyses of Variance (ANOVAs) were conducted on each set of content scales for the three groups by sex. Group A was presented first with the Strong, second with the MPQ, and third with the

ESCI. Group B was presented first with the ESCI, second with the MPQ, and third with the Strong. Group C was presented first with the MPQ, second with the Strong, and third with the ESCI. One-way ANOVAs were conducted for Groups A, B, and C on the 12 GOTs, 55 BISs, and 9 PSSs of the1994 and 2004 Strongs by sex.

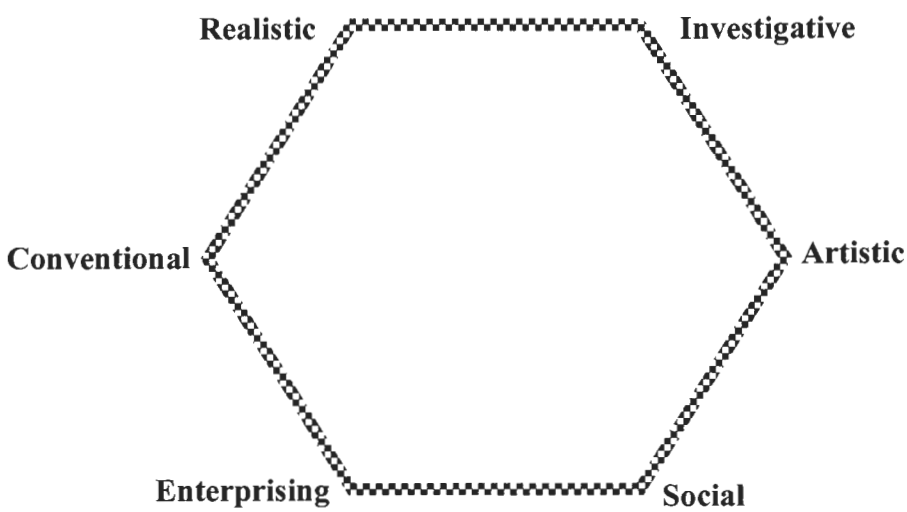
Each content scale was considered individually by sex, and for each of the 152 comparisons, a Bonferroni adjusted p-value less than .0003 was used as the criterion to determine statistical significance. Using this criterion, only four analyses yielded significant results. All four significant content scales were for women only and occurred on the 2004 Strong. There was one 2004 GOT (Conventional) and three 2004 BISs falling under the Conventional GOT that were significant ($p < .0003$) for women. Specifically, the following three BISs were significant: 1) Sales, $F(2, 368) = 9.26$, $p < .0003$, 2) Management, $F(2, 368) = 9.15$, $p < .0003$, and 3) Office Management, $F(2, 368) = 9.99$, $p < .0003$. Post hoc comparisons revealed that women in Group A (those completing the Strong first) scored significantly lower than women in Group B (those completing the Strong last) on these four scales. No comparisons with Group C were significant. However, the greatest mean difference observed was 5.74 points (95% CI = 2.53 to 8.96) for the Sales BIS which approximates one-half of a standard deviation. When combined with the low percentage of significant scales (4/152 = 2.63%) it appears that order exerted only a nominal effect.

1994 Strong Interest Inventory

The Strong is a widely used vocational assessment instrument designed to assess individual interests relative to occupations, people, environments, and leisure activities (Hansen & Campbell, 1985). The Strong incorporates Holland's theory for describing interests and their relationship to jobs, people, and environments (Hansen, 1994), thus

providing theoretical structure. In his theory, Holland (1985) divides both people and environments into some combination of six vocational personality types. These six types are hexagonally organized and include: the Realistic type (outdoors, mechanical), the Investigative type (science, math), the Artistic type (art, language), the Social type (helping, teaching), the Enterprising type (selling, business) and the Conventional type (details, clerical) (Hansen, 1995). Figure 1 below provides a graphic representation of the six Holland types arranged hexagonally.

Figure 1. A Diagrammatic Representation of the Six Holland Types (RIASEC).



*Note: Adapted From Strong Interest Inventory: Applications and Technical Guide (*Harmon, Hanson, Borgen, & Hammer, 1994).

The Strong includes over 100 different measures of interests organized around these six Holland codes, allowing it to offer global measures of interest via the scores obtained on the GOT scales (Hanson, 1985). The Strong also offers respondents three additional types of information: first, scores on the 25 BISs provide information about interests in specific areas; second, scores on the 211 OSs (which represent 109 different occupations) provide

information about interest similarity between the respondent's interests and those currently working in the relevant field; and third scores on the four Personality Scales provide information about the style with which the respondent prefers to learn, work, lead, and take risks (Harmon et al., 1994). There are also three types of Administrative (or validity) Indices on the 1994 Strong, but as with the OSs, they are not included in the scope of this study. Table 2.1 denotes the 1994 Strong BIS and PSS scales with descriptors.

Table 2.1

1994 Strong BIS and PSS Scales with Descriptors

GOT (in bold) & BIS	DESCRIPTOR
Realistic	
Agriculture	Interest in farming, physically active work in outdoor settings
Nature	Interest in outdoor recreational activities, appreciation for nature
Military Activities	Interest in well-ordered structured environments
Athletics	Intense interest in sports either as a participant or fan
Mechanical Activities	Interest in working with large equipment and precise instruments
Investigative	
Science	Interest in the natural sciences and scientific research
Mathematics	Interest in working with numbers and statistical analyses
Medical Science	Interest in the biological sciences and medical fields
Artistic	
Music/Dramatics	Interest in the performing/enjoying music or drama
Art	Interest in observing and enjoying works of art
Applied Arts	Interest in participating in varied performance activities
Writing	Interest in writing and reading literary works
Culinary Arts	Interest in cooking and entertaining

Table 2.1 (Continued)

Social	
Teaching	Interest in teaching young people w/ lots of social interaction
Social Service	Humanistic interest in working with and helping people
Medical Service	Interest in providing direct service to patients in medical settings
Religious Activities	Interest in spiritual or religious concerns
Enterprising	
Public Speaking	Interest in persuading and influencing others verbally
Law/Politics	Interest in debates and arguments designed to sell concepts
Merchandising	Interest in wholesale and retail activities designed to sell goods
Sales	Interests in calling on new customers to make a sale
Organizational Management	Interest in authority and power and in supervising and organizing
Conventional	
Data management	Interest in using data and information to make decisions
Computer Activities	Interest in working with computers and office machines
Office Services	Interest in clerical and office activities

PSS	**DESCRIPTOR**
Work Style Scale	Measures aspects of the style in which an individual likes to work (i.e., with ideas, data, things vs. with people)
Learning Environment Scale	Measures aspects of the style in which an individual likes to learn (i.e., in an academic vs. a hands-on environment)
Leadership Style Scale	Measures aspects of the leadership style an individual prefers (i.e., prefers to lead vs. prefers to follow)
Risk Taking/Adventure Scale	Measures aspects of the individual's risk-taking style (i.e., prefers adventure and risk vs. prefers stability and safety)

The 1994 Strong (Strong; Harmon et al., 1994) consists of 317 items, which comprise the scales mentioned above. It is preceded by the Strong Vocational Interest Blank (Strong,

1927), and by revisions completed in 1974 and 1985 (Campbell, 1974; Hansen & Campbell, 1985). Of the 18,951 people who made up the General Reference Sample of the 1994 Strong, 9,647 were women and 9,484 were men who represented relatively equally each of the 98 occupational samples collected for this revision (Harmon et al.). The vast majority of the items on each of the scales of the 1994 Strong are unit-weighted, so that high scores are related to high levels of interests. The three exceptions to this are the OSs (not included in this study), the Work Style Personality scale, and the Learning Environment Personality Style scale, which were developed by the contrasted-groups method with a significant proportion of items negatively weighted. The actual scoring weights for these measures are proprietary, so they are not released to the public. The Harmon et al. manual (1994) shows sample items for each scale to convey its content, but other scoring methods are necessarily excluded.

General Occupational Themes (GOTs). The GOTs are the broadest content scales on the 1994 Strong. Often, content scales are specifically constructed to fit a theoretical or conceptual map of important and independent interest domains (Borgen, in press). The GOTs are intended to align with Holland's six RIASEC code types. Harmon et al. (1994) reported Cronbach alphas for each of the GOTs as .90 or higher, which supports the RIASEC configuration proposed by Holland and demonstrates the GOT's homogeneity.

Additionally, Harmon et al. (1994) used test-retest reliability analyses with four separate samples to determine the stability of the GOTs. A list of all 24 test-retest correlations is presented in chapter five of the *Strong Interest Inventory Applications and Technical Guide* (p. 57), but for purposes of brevity only the lowest correlation and the highest correlation, representing the range, are presented here. The lowest correlation, $r = .4$,

occurred in the fourth sample (one of three college student examples with the fourth sample being one of working adults) for the Enterprising theme. The highest correlation, $r = .92$, was achieved in the single employed adults sample for the Realistic theme.

Speaking to the validity of the GOTs, Tracey & Rounds (1993) analysis showed that the GOTs provide the best-published inventory measures of the RIASEC dimensions. The concurrent validity of the GOTs is high as tested against the Vocational Preference Inventory (Hansen, 1986) with a .765 median correlation between the same-named scales of both inventories.

Basic Interest Scales (BISs). The Basic Interest Scales are further sub-dividers of the GOTs, and provide more specific measurement of the participants' homogenous interest areas (Donnay & Borgen, 1996). In this revision, two scales were removed (Adventure and Domestic Arts), four new scales were added (Applied Arts, Culinary Arts, Data Management, and Computer Activities), and the item content of 16 additional scales was revised (Harmon et al., 1994).

Both the reliability and validity of the BISs has been well established. The internal consistency and homogeneity of item content on the 25 BISs of the 1994 Strong are high, with Cronbach alphas ranging from .74 for the Agriculture scale (6-items), to .94 for the Mechanical Activities scale (21-items) (Harmon et al., 1994). In addition, these same researchers reported three to six month test-retest correlations for the BISs ranging from .80 for Culinary Arts and Teaching, to .94 for Athletics. The face validity of the BISs is also high with item content being clearly discernable by the name of each scale (i.e., Solving Mechanical Puzzles). Finally, with increased specificity comes increased predictive ability,

and research has shown that the BISs are generally more predictive of occupation than are the GOTs (Hansen, 1986). Refer to Table 2 for specific descriptors associated with each BIS.

Personal Style Scales (PSS). The 1994 Strong introduced the PSSs to the world, replacing the Special Scales of the 1985 Strong Interest Inventory. Again, the four PSSs include Work Style, Learning Environment, Leadership Style, and Risk Taking/Adventure (refer to Table 2). Two of the PSSs (Work Style and Learning Environment) were empirically derived from two groups of participants; the first of whom preferred working with people (n = 6,681) and the second of whom preferred working with data, ideas, or things (n = 5,574). From these participants' responses, 51 items were identified which successfully differentiated between the two groups (Harmon et al., 1994). For the Learning Environment Scale, again two groups were tested. The first group consisted of participants who had earned graduate degrees, either their Master's degree or their Ph.D. (n = 7,328), and the second group consisted of participants who had completed either a trade or a technical course (n = 422). The 49 resulting items that successfully discriminated between the two groups were selected to comprise the Learning Environment Scale (Harmon et al.). A third scale, the Risk Taking/Adventure scale, was first developed by Campbell, Borgen, Eastes, Johansson, and Peterson in 1968. The 1994 version of this scale retains the nine identical items of the 1985 revision, with only a few reworded to reflect more contemporary language. The construction of the final scale, Leadership Style, was accomplished by the use of a factor analysis to select highly inter-correlated items (Harmon et al.).

The PSS are moderately inter-correlated with the highest correlation falling between the Leadership Style and Work Style scale (rs = .61, .52 for men and women respectively), and Work Style and Leaning Environment showing nominal correlations (Harmon et al.,

1994). An evaluation of the internal consistency of the PSSs resulted in Cronbach alphas of .91 for Work Style, .86 for Learning Environment, .86 for Leadership Style, and .78 for Risk Taking/Adventure (Harmon et al.). Three to six month test-retest reliabilities were examined using the same four samples (three college and one working adult) described previously and ranged from a high of .92 for the Work Style scale for the working adult sample, to a low of .81 for the Leadership Style scale from the third college sample (Harmon et al.). The validity of the PSSs is strongly supported by their clear ability to differentiate (more than one standard deviation) between educational majors (Harmon et al.). Again, the OSs were not examined in this study.

2004 Strong Interest Inventory

The 2004 Strong contains 291-items and three types of content scales: six General Occupational Themes, 30 Basic Interest Scales, and five Personality Style Scales. The original intent of the revisers of the 2004 Strong (Donnay et al., 2005) was to replicate the data collection procedures used for the 1994 revision to create the new General Representative Sample (GRS). Towards this goal, 73,500 potential respondents were invited to participate via mass mailing procedures between 2001 and 2002 (Donnay et al.). Unfortunately, response rates were unusually low. The developers thus decided to use a different approach reflective of the more technologically modern culture preferring to complete the Strong online.

Specifically, two data collection approaches were utilized. First, to garner participants in specific occupations needed for the revision of the occupational scales, organizations and associations containing individuals engaged in specific occupations were contacted directly. Second, respondents for the GRS were garnered via advertisements placed on Internet search

engine sites. Those individuals identified as searching for the term "career" were targeted and presented with an advertisement inviting them to participate in a free career assessment. In total, over 20,000 individuals responded to the advertisement and completed the research version of the Strong online. However, due to the randomness of such a sample, most of the respondents did not meet the criteria necessary for inclusion in the GRS and in particular in the occupational sample (e.g., too young, too few years in their chosen occupation).

In total, 2,250 respondents (mean age = 35 years) met the criteria and were chosen to be included in the GRS with half being women and half being men. The ethnicity for females and males closely approximated one another and sufficiently represented the major racial and ethnic groups in the United States (Donnay et al., 2005). The GRS also adequately represented the world of work with its sample of working adults encompassing more than 370 occupations. Additionally, GRS respondents reported working an average of 41 hours per week in their respective occupations. Finally, an average of more than nine years of experience working in the current occupation was reported. Table 2.2 below provides a breakdown of ethnicity by sex for the 2004 GRS.

Table 2.2

2004 Strong Interest Inventory GRS Distribution of Race/Ethnicity by Sex

Race/Ethnicity	**Women (*n* = 1,125)**	**Men (*n* = 1,125)**
Native American	.6%	.1%
Alaskan Native	.0%	.2%
Hawaiian Native or other Pacific Islander	.4%	.3%

Table 2.2 (Continued)	**Women (*n* = 1,125)**	**Men (*n* = 1,125)**
Asian American or Asian	4.3%	2.2%
Caucasian/White	68.3%	76.2%
Asian Indian	1.6%	2.4%
Latino, Latina/Hispanic	6.2%	3.8%
Multiethnic	6.4%	6.5%
No Response	1.5%	2.3%

Note. N = 2,250. Adapted from Strong Interest Inventory manual: Research, development, and strategies for interpretation (Donnay et al., 2005).

Of note is that participants included in the norming of the new 2004 Occupational Scales all met the following criteria: they were at least 25 years of age, they had at least three years of relevant job experience, they performed activities typical for their occupation, and they reported overall satisfaction within their occupation. Additionally, though the OSs are not included in the current study, it is important to note that the 2004 OSs retained from the 1994 Strong were derived using archival databases; while the new 2004 OSs were derived from 2,817 (1,537 women and 1,280 men) who completed the research form of the 2004 Strong. In total, there were 36 new occupations originally conceptualized for inclusion in the new Strong. Of these, 13 occupations were garnered through traditional sampling methods used in past versions and 11 were obtained online (Donnay et al., 2005). Thus, the GRS and the OS samples are not equivalent.

The 2004 Strong was designed to retain most of the structural integrity of its 1994 predecessor, therefore much of the information related to the content scales of the 1994

Strong remains valid for the 2004 version. However, some changes were naturally made to each set of scales. Table 2.3 presents descriptions of the 2004 BISs and PSSs.

Table 2.3

2004 Strong BIS and PSS Scales with Descriptors

GOT (in bold) & BIS	DESCRIPTOR
Realistic	
Mechanics & Construction	Interest in working with large machinery, precise instruments
Computer Hardware & Electronics	Interest in installing and repairing computer hardware and networking systems
Military	Interest in well-ordered structured environments
Protective Services	Interest in nonmilitary-related public safety and policing
Nature & Agriculture	Interest in working in farming or ranching, enjoying nature
Athletics	Interest in sports (particularly team sports)
Investigative	
Science	Interest in the natural sciences (especially physical science)
Research	Interest in designing and conducting studies
Medical Science	Interest in the biological sciences and medical fields
Mathematics	Interest in working with numbers and statistical analyses
Artistic	
Visual Arts & Design	Interest in visual creativity and spatial visualization
Performing Arts	Interest in participating in varied performance activities
Writing & Mass Communication	Interest in literature, reading, and language (both appreciating and producing)
Culinary Arts	Interest in cooking and entertaining
Social	
Counseling & Helping	Altruistic interest in working with and helping people
Teaching & Education	Interest in the highly-interactive teaching of young people
Human Resources & Training	Interest in developing and training people and managing and directing organizational activities
Social Sciences	Interest in the study of people, groups, society, and cultures
Religion & Spirituality	Interest in religious or spiritual concerns
Healthcare Services	Interest in providing service in medical settings

Table 2.3 (Continued)

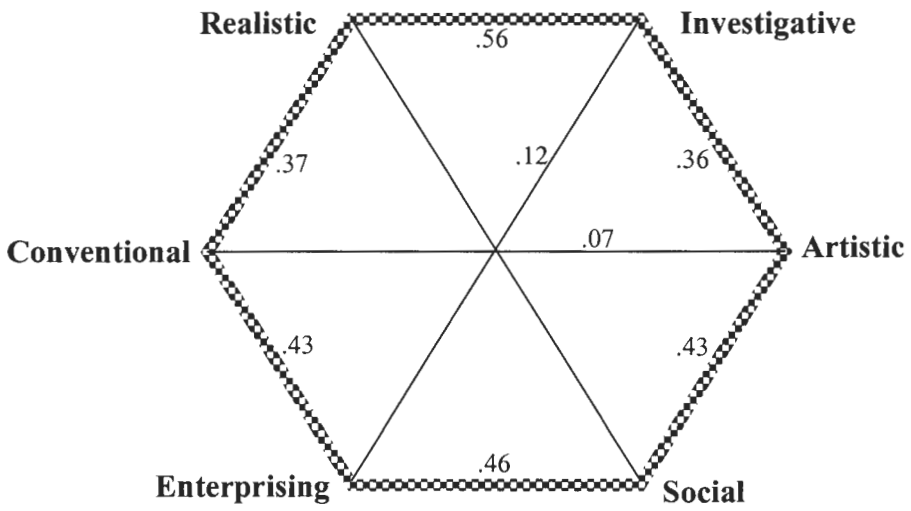
Enterprising	
Public Speaking	Interest in persuading and influencing others verbally
Law/Politics	Interest in debates and arguments designed to sell concepts
Merchandising	Interest in wholesale/retail activities designed to sell goods
Sales	Interests in calling on new customers to make a sale
Organizational Mngmt	Interest in authority, power, supervising, and organizing
Conventional	
Data management	Interest in using data and information to make decisions
Computer Activities	Interest in working with computers and office machines
Office Services	Interest in clerical and office activities

PSS	DESCRIPTOR
Work Style Scale	Measures aspects of the style in which an individual likes to work (i.e., with ideas, data, things vs. with people)
Learning Environment Scale	Measures aspects of the style in which an individual likes to learn (i.e., in an academic vs. a hands-on environment)
Leadership Style Scale	Measures aspects of the leadership style an individual prefers (i.e., prefers to lead vs. prefers to follow)
Risk Taking/Adventure Scale	Measures aspects of the individual's risk-taking style (i.e., prefers adventure and risk vs. prefers stability and safety)

General Occupational Themes (GOTs). To retain its interpretative validity, the developers of the new Strong retained the hexagonal structure of the GOTs based on Holland's theory (1985). Moreover, any differences observed in the 2004 GOTs most likely reflect changes made to the BISs. The main effect of the BIS changes was to broaden and expand the GOTs to include technological advances in the work place. For example, the Conventional theme was broadened and now includes items measuring computer programming and working with software (Donnay et al., 2005). This expansion of the GOTs

served to increase the internal consistency reliability of five of the themes (Realistic remained constant at .93), with all six themes possessing Cronbach alpha's of .91 or higher (Donnay et al.). The inter-scale correlations also remained consistent as shown in Figure 2 below.

Figure 2. 2004 Intercorrelations of the Six Holland Types (RIASEC).



Note: Adapted From Strong Interest Inventory Manual: Research, development, and strategies for interpretation (Donnay et al., 2005)

Basic Interest Scales (BISs). The BISs were increased from 25 in the 1994 Strong to 30 in the 2004 Strong. In all, four scales were dropped and ten new scales were added, with the remaining BISs being revised and updated (Donnay et al., 2005). Additionally, the number of items per scale was reduced from 5-21 items per BIS in 1994 to 6-12 items per BIS in 2004 (Donnay et al.). The overall internal consistency remained unchanged for the 2004 Strong BISs with a median Cronbach's alpha of .87. Table 2.4 presents a summary of the additions and changes made to the 2004 Strong BISs.

Table 2.4

Summary of Changes to the *2004 Strong Interest Inventory Basic Interest Scales*

		Change to 2004 BIS				
2004 BIS	**1994 BIS**	**Added**	**Changed**	**Merged**	**Divided**	**n/a**
Mechanics & Construction	Mechanical Activities		X			
Comp. Hardware & Electronics	N/A	X				
Military	Military Activities		X			
Protective Services	N/A	X				
Nature & Agriculture	Nature/Agriculture			X		
Athletics	Athletics					X
Science	Science					X
Research	N/A	X				
Med. Science	Med. Science					X
Mathematics	Mathematics					X
Visual Arts & Design	Applied Arts		X			
Performing Arts	Music/Dramatics		X			
Writing & Mass Communication	Writing		X			
Culinary Arts	Culinary Arts					X

Table 2.4 (Continued)

Counseling & Helping	Social Service		X			
Teaching & Education	Teaching		X			
Human Resources & Training	N/A	X				
Social Sciences	N/A	X				
Religion & Spirituality	Religious Activities		X			
Healthcare Services	Medical Service		X			
Marketing & Advertising	N/A	X				
Sales	Sales					X
Management	Organizational Management		X			
Entrepreneurship	N/A	X				
Law	Law/Politics		X	X	X	
Office Management	Office Services		X			
Taxes & Accounting	N/A	X				
Programming & Information Systems	N/A	X				
Finance & Investing	N/A	X				

Note: Adapted From Strong Interest Inventory Manual: Research, development, and strategies for interpretation (Donnay et al., 2005)

Personal Style Scales (PSSs). The four Personal Style Scales first introduced in the 1994 Strong remained relatively unchanged in the 2004 version, with the exception of the Risk Taking scale which was more substantially updated to reflect a wider array of risk taking behaviors (e.g., emotional, financial, and physical risks). However, a new PSS, Team Orientation, was added to the 2004 bringing its total PSSs to five. As with the other PSSs, Team Orientation is measured on two poles. Its two poles range from preferring to accomplish tasks independently to preferring to accomplish tasks collectively (Donnay et al., 2005). The PSSs, like the GOTs and BISs, were normed on the GRS (n = 2,250). The inter-correlations between scales was acceptable and ranged from .03 between the Work Style scale and the Leadership scale to .55 between the Risk Taking and Team Orientation scale (Donnay et al.).

Preliminary Analyses

This study sought to help aid in the user acceptance of the new 2004 Strong by establishing its relative equivalence to the 1994 version. Specifically, the author conducted the following three preliminary analyses. 1) It was expected that independent sample t-tests would show no meaningful differences in the means and standard deviations of the content scales between the current sample, the 2004 GRS, and the 2005 Gasser sample. 2) It was expected that Shapiro-Wilks's tests of normality would show no meaningful differences between the content scales of the 1994 and 2004 Strong Interest Inventories. 3) It was expected that univariate analyses of variance (ANOVAs) on the content scales of the 1994 and 2004 Strong Interest Inventories would demonstrate equivalent statistically significant differences between sexes.

Independent sample t-tests were used to test the first preliminary analysis that no meaningful differences would exist in the means and standard deviations of the content scales between the current sample, the 2004 GRS, and the 2005 Gasser sample. This first preliminary analysis also addressed the representativeness of the current sample. The second preliminary analysis used Shapiro-Wilks's tests of normality to show that no meaningful differences would exist between the content scales of the 1994 and 2004 Strong Interest Inventories regarding normal distribution. Univariate analyses of variance (ANOVAs) were used to test the third preliminary analysis that the content scales of the 1994 and 2004 Strong Interest Inventories would demonstrate equivalent statistically significant differences between sexes. Specifically, the ANOVAs were used to determine the existence and magnitude of sex effects for each of the content scales of the 1994 and 2004 Strongs.

Hypotheses

Beyond the aforementioned preliminary analyses, the author tested the following two hypotheses. 1) It was expected that paired sample t-tests would show no meaningful differences between the means of the parallel content scales of the 1994 and 2004 Strongs. 2) It was additionally expected that strong positive relations would be present between the analogous content scales of the 1994 and 2004 Strong Interest Inventories.

Analyses

The first hypothesis was tested using paired samples t-tests. Specifically, paired samples t-tests were used to test the first main hypothesis that no meaningful differences would exist between the means of the parallel content scales of the 1994 and 2004 Strong Interest Inventories.

Bivariate correlational analyses were conducted to test this second hypothesis that no meaningful differences would exist between parallel content scales of the 1994 and 2004 Strong Interest Inventories. These Pearson product moment correlations were used to test the equivalency of participant scores on the analogous content scales (GOTs, BISs, PSSs) of the 2004 and 1994 Strongs.

CHAPTER 3. RESULTS

This section is divided into four parts. It begins with a discussion of the results of the preliminary analyses mentioned earlier. This will be followed by a discussion of the differences between the means and standard deviations of the parallel content scales of the 1994 and 2004 Strong Interest Inventories presented by sex, per the first hypothesis. Finally, the bivariate relationships between the analogous content scales of the 1994 and 2004 Strongs will also be discussed by sex, which addresses the second hypothesis.

Preliminary Analyses

Three preliminary analyses will be discussed in this section. The first analysis presented examines the representativeness of the current sample by using independent samples t-tests to compare the 2004 content scales of the current sample to both the 2004 GRS and to the Gasser sample (Gasser, 2005) by sex. Next, the second analysis presents the Shapiro-Wilks tests of normality of the distributions of the current sample for each of the 35 content scales of the 1994 Strong and for each of the 41 content scales of the 2004 Strong by sex. Finally, the third analysis presents the univariate analyses of variance (ANOVAs) used to test the equivalency of the statistically significant differences between sexes.

Comparison of 2004 SII Scales with the General Reference Sample and the Gasser Sample by Sex

To determine the representativeness of the current sample, descriptive analyses were conducted. That is, independent samples t-tests were run comparing the current sample of college students to the established norms of the 2004 Strong General Reference Sample (i.e., a sample of employed adults) and to the 2004 Gasser sample (i.e., a comparison sample of college students). In all, 82 comparisons were run for both females and males. Due to the

high number of comparisons, a Bonferroni adjustment was used (.05 / 164 comparisons = .0003; $t = 3.62$).

General Occupational Themes. In Table 3.1, the GOTs in the current sample were compared to the GOTs in both the GRS and in the Gasser sample. As illustrated by the table, only one GOT, Investigative, did not differ significantly by sex for either the GRS or for the Gasser sample. Two of the GOTs, Artistic and Conventional, displayed sex differences with only the Artistic females differing significantly from the Gasser sample (they did not differ from the GRS), and only the Conventional males failing to differ significantly from the GRS (all other Conventional comparisons differed significantly). For the Realistic, Social, and Enterprising scales, all mean differences were statistically different between samples for both sexes ($p < .0003$).

The implications of these mean differences are meaningful but not alarming. No mean differences between the current sample and the GRS differed by more than one-half of a standard deviation. Interestingly, for males only, Two GOT mean differences between the current sample and the Gasser sample differed by more than one-half of a standard deviation, but by less than a full standard deviation (i.e., a mean difference of 7.69 for the male Realistic scale and a mean difference of 6.07 for the male Conventional scale; 1 $SD = 10$).

Additionally, the eta squared values were calculated for each scale to determine the effect size. Eta squared falls on a continuum from 0 to 1 and represents the proportion of variance in the dependent variable (i.e., GOT scale) that is explained by the group variable (i.e., the 1994 versus the 2004 Strong). Using Cohen's (1988) guidelines, for interpreting the eta values, all but three of the observed mean differences had small effect sizes (eta squared = .01 to .04). Two GOTs in the Gasser sample showed a medium effect size of .07 (female

Realistic and male Conventional). Only one GOT garnered a large effect size and that was the male Realistic scale between the current sample and the Gasser sample (eta squared = .14).

Table 3.1

Means and Standard Deviations for 2004 Strong Interest Inventory General Occupational Themes—Comparison of Current Sample with 2004 GRS and 2005 Gasser Sample

	Females						**Males**					
	GRS (n = 1125)		Gasser (n = 1403)		Current (n = 374)		GRS (n = 1125)		Gasser (n = 470)		Current (n = 248)	
	Mean	*SD*	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
R	44.97	8.42	42.40	7.92	***47.52***	***8.01***	55.03	8.87	50.75	9.65	***58.44***	***7.98***
I	48.56	10.12	47.62	10.54	48.24	9.51	51.44	9.67	49.28	10.82	52.18	9.18
A	51.31	10.19	50.15	10.18	*52.57*	*8.83*	48.69	9.64	46.72	9.96	48.37	8.92
S	51.94	9.92	52.10	10.44	***56.58***	***9.59***	48.06	9.71	46.30	11.21	***50.49***	***8.61***
E	49.61	9.81	47.67	10.84	***52.80***	***9.64***	50.39	10.18	49.70	11.54	***53.36***	***9.50***
C	49.43	10.63	46.66	10.75	*51.41*	*10.82*	50.57	9.30	49.66	10.53	***55.73***	***10.34***

Note. RIASEC = Realistic, Investigative, Artistic, Social, Enterprising, Conventional. Bold values denote statistically significant difference from the 2004 GRS. Italicized values denote statistically significant difference from the 2005 Gasser sample. Bold and italicized values

denote statistically significant difference from both the 2004 GRS and the 2005 Gasser sample.

Basic Interest Scales. Table 3.2 illustrates the differences between the means of the Basic Interest Scales for the current sample and both the GRS and the Gasser sample. Only seven of the BISs between the current sample and the GRS were statistically significant ($p <$ .0003) for females. These differences included the following scales: Protective Services, Athletics, Medical Science, Teaching and Education, Healthcare Services, Marketing and Advertising, and Sales. However, over twice as many statistically significant mean differences were found between the current sample and the Gasser sample with exactly half ($n = 15$) of the BISs differing for females. Again, these mean differences included the following scales: Mechanics and Construction, Military, Protective Services, Nature and Agriculture, Athletics, Medical Science, Visual Arts and Design, Performing Arts, Teaching and Education, Healthcare Services, Marketing and Advertising, Sales, Management, Office Management, and Taxes and Accounting.

As before, the meaning of these mean differences can be gauged by looking at how large the differences are between the means. The majority of the differences between the means ($n = 16$) were less than one-half of a standard deviation. Those mean differences between the current sample and the comparison samples over one-half a standard deviation included the following: Athletics (GRS = 8.44; Gasser = 7.85), Teaching and Education (GRS = 6.06), Healthcare Services (GRS = 5.91; Gasser = 6.02), and Sales (Gasser = 6.25). As with the GOTs, the eta squared values were computed for each difference to calculate the magnitude of its effect size and reflect the proportion of variance in the dependent variable (i.e., BIS scale) that is explained by the group variable (i.e., the 1994 versus the 2004

Strong). Eighty-one percent (n = 17 out of 21) of the statistically significant differences in the means in the BISs had a small effect size (eta squared = .01 to .05). There were three scales with medium effect sizes including the following: Athletics (Gasser sample; eta squared = .10), Teaching and Education (GRS; eta squared = .06), and Sales (Gasser sample; eta squared = .06). Only a single BIS displayed a large effect size; the Athletics scale between the current sample and the GRS had an eta squared value of .14.

Slightly more significant differences between the means of the current sample and the GRS and Gasser Sample were found for males. Specifically, there were 12 statistically significant differences ($p < .0003$) between the current sample and the GRS including the following: Mechanics and Construction, Military, Protective Services, Athletics, Medical Science, Math, Teaching and Education, Healthcare Services, Sales, Management, Office Management, and Taxes and Accounting. As with the females, a greater number of differences were found between the current sample and the Gasser sample. That is, 16 BISs differed significantly ($p < .0003$) and included the following: Mechanics and Construction, Computer Hardware and Electronics, Military, Protective Services, Nature and Agriculture, Athletics, Science, Medical Science, Math, Visual Arts, Teaching and Education, Healthcare Services, Sales, Management, Office Management, and Taxes and Accounting.

Unlike the females, only half of the differences between the means of the BISs for the males were less than one-half of a standard deviation (n = 14 out of 28 total differences across both samples). The greatest difference occurred between the current sample and the Gasser sample on the Protective Services scale (i.e., 7.76). The eta squared values for the males also displayed an overall greater magnitude of difference between the means. Small effect sizes were found for the following scales: Mechanics and Construction (GRS; eta

squared = .02), Computer Hardware and Electronics (Gasser sample; eta squared = .03), Military (GRS; eta squared = .01), Medical Science (GRS; eta squared = .03), Math (GRS; eta squared = .02 and Gasser sample; eta squared = .04), Teaching and Education (GRS; eta squared = .02), Healthcare Services (GRS; eta squared = .05), Sales (GRS; eta squared = .05), Management (GRS; eta squared = .01 and Gasser sample; eta squared = .03), and for Office Management and for Taxes and Accounting (GRS; eta squared = .04 and .03, respectively). Medium effect sizes were found for the following scales: Military (Gasser sample; eta squared = .08), Protective Services (GRS; eta squared = .06 and Gasser sample; eta squared = .12), Nature and Agriculture (Gasser sample, eta squared = .08), Athletics (GRS; eta squared = .06 and Gasser sample; eta squared = .07), Medial Science (Gasser sample; eta squared = .06), Teaching and Education (Gasser sample; eta squared = .06), Healthcare Services (Gasser sample; eta squared = .07), Sales (Gasser sample; eta squared = .08), Office Management (Gasser sample, eta squared = .06), and Taxes and Accounting (Gasser sample; eta squared = .09). Finally, only one BIS had a moderately large effect size, Mechanics and Construction (Gasser sample; eta squared = .13).

Table 3.2

Means and Standard Deviations for 2004 Strong Interest Inventory Basic Interest Scales—Comparison of Current Sample with 2004 GRS and 2005 Gasser Sample

	Females						**Males**					
	GRS		Gasser		Current		GRS		Gasser		Current	
	(n = 1125)		(n = 1403)		(n = 374)		(n = 1125)		(n = 470)		(n = 248)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
R												
Mechan & Constr	45.47	8.46	42.62	7.70	*46.16*	*7.74*	54.53	9.35	50.25	9.25	***57.76***	***8.95***

Table 3.2 (Continued)

	Females						Males					
	GRS (n = 1125)		Gasser (n = 1403)		Current (n = 374)		GRS (n = 1125)		Gasser (n = 470)		Current (n = 248)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Comput Hardwr & Elect	46.42	9.21	43.17	8.62	44.02	8.32	53.58	9.46	50.59	9.76	*53.76*	*8.91*
Military	46.11	8.25	44.97	8.54	*47.76*	*8.23*	53.89	10.08	50.56	10.16	***56.63***	***9.61***
Protect Services	47.29	9.56	46.89	9.49	***51.83***	***9.01***	52.71	9.69	51.20	10.50	***58.96***	***8.40***
Nature & Agricult	48.32	10.37	44.93	9.89	*49.92*	*9.12*	51.68	9.33	46.48	10.37	*52.48*	*8.19*
Athle-tics	47.12	9.09	47.71	9.67	***55.56***	***9.08***	52.88	10.05	53.22	10.56	***59.20***	***9.11***
I												
Science	47.95	9.93	46.70	10.27	47.97	9.26	52.05	9.64	49.84	10.61	*52.77*	*8.91*
Re-search	48.40	10.15	46.17	10.52	44.78	9.51	51.60	9.58	48.72	10.87	48.59	9.27
Medical Science	49.71	10.39	49.61	10.90	***53.68***	***10.63***	50.29	9.59	49.41	10.98	***54.88***	***9.23***
Mathe-matics	47.77	9.84	47.95	9.98	49.49	9.26	52.23	9.66	50.96	10.42	***55.33***	***9.15***
A												
Visual Arts & Design	50.74	10.37	48.51	10.20	*51.77*	*9.11*	49.26	9.56	46.65	9.65	*49.50*	*9.13*
Perf. Arts	51.82	10.01	51.20	10.63	*53.53*	*9.34*	48.18	9.66	47.64	10.27	47.82	9.99
Writing & Mass Comm	50.79	10.39	49.86	10.57	53.68	9.92	49.21	9.54	46.71	10.38	46.67	9.10

Table 3.2 (Continued)

	Females						Males					
	GRS (n = 1125)		Gasser (n = 1403)		Current (n = 374)		GRS (n = 1125)		Gasser (n = 470)		Current (n = 248)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Culinary Arts	51.57	9.69	51.14	9.75	52.95	8.44	48.43	10.06	47.66	10.92	48.71	8.82
S												
Counsel & Helping	52.29	9.88	52.59	10.36	53.97	8.93	47.71	9.59	46.54	11.02	47.99	8.28
Teach & Educ	50.81	10.32	51.87	11.14	***56.87***	***10.64***	49.19	9.61	47.28	11.04	***52.29***	***9.36***
Human Resrs & Trng	50.57	10.36	47.75	10.66	48.93	9.55	49.43	9.59	46.83	11.48	47.24	8.42
Social Science	50.42	10.24	51.24	10.01	50.38	8.19	49.58	9.74	49.28	10.70	49.63	8.18
Relig & Spirituality	50.23	9.79	49.23	10.14	50.81	9.61	49.77	10.20	47.90	10.43	49.72	9.45
Healthcare Services	51.18	10.73	51.07	10.89	***57.09***	***10.49***	48.82	9.07	48.30	10.43	***54.16***	***8.84***
E												
Market & Advert	51.03	10.13	49.16	11.16	***53.53***	***9.67***	48.97	9.76	48.76	11.29	51.41	9.30
Sales	48.82	9.29	47.74	9.67	***53.99***	***10.68***	51.18	10.53	50.74	11.52	***57.52***	***10.84***
Management	48.93	9.88	47.44	10.17	*49.88*	*9.63*	51.07	10.01	46.83	10.69	***53.76***	***9.54***
Entrepreneurship	48.94	10.05	44.85	11.15	45.97	9.64	51.06	9.85	48.81	11.19	48.76	9.77

Table 3.2 (Continued)

	Females						Males					
	GRS (n = 1125)		Gasser (n = 1403)		Current (n = 374)		GRS (n = 1125)		Gasser (n = 470)		Current (n = 248)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Politics & Pub. Speak	47.42	9.57	49.30	10.50	48.77	9.75	52.58	9.76	52.35	10.76	52.67	8.79
Law	49.51	10.06	49.70	10.78	50.64	10.16	50.49	9.92	50.62	10.62	52.87	9.21
C												
Office Management	52.27	10.77	49.98	10.38	*54.51*	*10.52*	47.73	8.59	47.74	9.18	***52.51***	***9.21***
Taxes & Acct	48.93	10.51	47.29	10.15	*50.55*	*9.98*	51.07	9.35	49.15	9.87	***55.57***	***9.48***
Prog & Info Sys	48.14	10.35	44.26	10.06	44.87	8.96	51.86	9.27	49.14	10.20	51.16	9.56
Finance & Invest	47.53	9.34	45.97	10.33	47.95	9.60	52.48	10.03	52.78	11.56	54.58	10.41

Note. RIASEC = Realistic, Investigative, Artistic, Social, Enterprising, Conventional. Bold values denote statistically significant difference from the 2004 GRS. Italicized values denote statistically significant difference from the 2005 Gasser sample. Bold and italicized values denote statistically significant difference from both the 2004 GRS and the 2005 Gasser sample.

Personal Style Scales. Table 3.3 denotes the statistically significant mean differences between the Personal Style Scales of the 2004 Strong between the current sample and both the 2004 GRS and the 2005 Gasser sample by sex. Only two of the five PSS means were statistically different between samples. The first difference occurred with the Work Style

scale, which also displayed a difference between sexes. For females, there was a statistically significant difference between the current sample and both the GRS and the Gasser Sample ($p < .0003$). However, for males, a statistically significant difference occurred only between the current sample and the GRS ($p < .0003$). The second difference occurred with the Risk Taking scale. For this scale there was a statistically significant difference between the means of the current sample and both the GRS and the Gasser samples for both females and males ($p < .0003$). Once again turning to Cohen's (1988) criteria for interpreting the effect size of the eta squared values, the meaning of these differences can be determined. For the Work Style scale, all effect sizes were small for both females and males (eta squared = .01 for males; .05 and .03 for females, GRS and Gasser samples respectively). The Risk Taking Scale displayed both greater disparity and magnitude of effect size with females showing moderate effects (eta squared = .07 for both the GRS and Gasser samples) and males showing a small effect between the current sample and the GRS (eta squared = .04), but a moderate effect between the current sample and the Gasser sample (eta squared = .07).

Table 3.3

Means and Standard Deviations for 2004 Strong Interest Inventory Personal Style Scales—Comparison of Current Sample with 2004 GRS and 2005 Gasser Sample

	Females						**Males**					
	GRS (n = 1125)		Gasser (n = 1403)		Current (n = 374)		GRS (n = 1125)		Gasser (n = 470)		Current (n = 248)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Work Style	54.16	9.17	55.16	9.44	***59.08***	***8.77***	45.84	9.02	46.71	9.41	**48.18**	**8.12**
Learning Environ	49.15	10.57	49.82	9.89	44.65	8.42	50.85	9.32	48.71	9.81	44.48	7.40

Table 3.3 (Continued)

	Females						Males					
	GRS (n = 1125)		Gasser (n = 1403)		Current (n = 374)		GRS (n = 1125)		Gasser (n = 470)		Current (n = 248)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Leader-ship	49.13	10.19	49.51	10.78	48.57	9.80	50.87	9.74	49.79	11.16	49.20	8.98
Risk Taking	45.49	8.82	45.32	8.93	***51.17***	***8.52***	54.51	9.03	53.53	10.21	***59.28***	***8.61***
Team Orient-ation	49.60	10.48	46.59	11.63	48.96	9.89	50.40	9.49	47.88	11.83	48.24	8.88

Note. Bold values denote statistically significant difference from the 2004 GRS. Bold and italicized values denote statistically significant difference from both the 2004 GRS and the 2005 Gasser sample.

Summary of Between Sample Comparisons Results. In total, 164 t-tests were run in an effort to determine the degree of similarity between the current sample and both the 2004 GRS and the 2005 Gasser sample. After making a Bonferroni adjustment, 73 of these comparisons were determined to be statistically significant at the .0003 level. However, due to the relatively large sample sizes of the three comparison samples, a statistically significant difference does not allow for direct interpretation of the meaningfulness of the observed differences. To determine this, eta squared values were calculated for each statistically significant difference to more accurately capture its magnitude. Based on Cohen's (1988) criteria, of the 73 comparisons, 69% showed small effect sizes (n = 50), 27% showed moderate effect sizes (n = 20), and 4% showed large effect sizes (n = 3). This relatively large percentage of small effect sizes, in combination with the large number of non-significant t-

tests (n = 91), suggests that the current sample is an adequate representation of both the GRS and the Gasser sample as suggested by the first preliminary analysis.

Tests of Normalcy between the 1994 and 2004 Strong Interest Inventories

The assumption of normality plays a pivotal role in statistical analyses as it is the basis for many statistical tests. More precisely, a normal distribution is considered a reasonable model of the expected behavior pattern of certain random phenomena, and as such can be used to approximate other probability distributions. One of the most highly recommended methods for testing the normality of data is the Shapiro-Wilks test. According to its developers (Shapiro & Wilks, 1965), this test rests on the premise that for normally distributed data, the ordered values should be highly correlated with corresponding quantiles (z-scores) taken from a normal distribution. This test was performed on each of the 35 content scales of the 1994 Strong and on each of the 41 content scales of the 2004 Strong by sex in the current sample. The null hypothesis for all of these tests is that the residuals are normally distributed. The results of each set of content scales are described in turn.

General Occupational Themes. Table 3.4 presents the tests of normality for the GOTs for the 1994 and 2004 Strongs by sex. The Shapiro-Wilks statistic (W; ranging from zero to one) presented in the table is essentially the ratio of the best estimator of the variance to the usual corrected sum of squares estimator of the variance. The significance level (p-value) of W is also presented with values greater then .002 (representing a Bonferroni adjustment for multiple tests .05 / 24 = .002) indicating normality. As can be seen from the table, there is a noticeable difference between the 1994 and 2004 Strongs, as well as a difference between females and males for the 1994 Strong. Specifically, seven of the 1994 GOT tests garnered significant results ($p < .002$) indicating a violation of the normality assumption. All of these

scales except for one (male Artistic) were significant only for the females, who showed statistical significance across all six GOTs. The remaining five male GOTs (i.e., Realistic, Investigative, Social, Enterprising, and Conventional) were not statistically significant denoting normal distribution. However, all 2004 GOT tests were non-significant ($p > .002$) for both sexes indicating normal distribution for all scales.

Table 3.4

Shapiro-Wilks's Tests of Normality for the General Occupational Themes of the 1994 and 2004 Strong Interest Inventories by Sex

		Shapiro-Wilks Tests		
GOT	**Sex**	**Statistic (W)**	**df**	***p*-value**
1994 Realistic	Male	.983	248	.004
	Female	.884	374	*<.001*
2004 Realistic	Male	.987	248	.027
	Female	.991	374	.027
1994 Investigative	Male	.985	248	.009
	Female	.956	374	*<.001*
2004 Investigative	Male	.997	248	.943
	Female	.990	374	.009
1994 Artistic	Male	.972	248	*<.001*
	Female	.984	374	*<.001*
2004 Artistic	Male	.992	248	.237
	Female	.990	374	.011

Table 3.4 (Continued)

		Shapiro-Wilks Tests		
GOT	**Sex**	**Statistic (W)**	**df**	***p*-value**
1994 Social	Male	.981	248	*.002*
	Female	.970	374	*<.001*
2004 Social	Male	.986	248	.019
	Female	.989	374	.008
1994 Enterprising	Male	.983	248	.005
	Female	.985	374	*.001*
2004 Enterprising	Male	.994	248	.431
	Female	.992	374	.053
1994 Conventional	Male	.980	248	*.002*
	Female	.969	374	*<.001*
2004 Conventional	Male	.994	248	.426
	Female	.993	374	.075

Note. All p-values in italics denote statistical significance at the .002 level

Basic Interest Scales. Table 3.5 presents the normality tests for the 1994 and 2004 BISs by sex. Not surprising giving the hierarchical structure of the Strong, the results of the BIS tests closely mimic and expound those of the GOTs. Once again, a Bonferroni adjustment (.05 / 110 = .0005) was made to the significance level (*p*-value) of W, thus values greater then .0005 indicates normal distribution. As with the GOTs, there is a noticeable difference between the 1994 and 2004 Strongs, as well as a difference between females and

males for both test versions. The results of for the 1994 Strong will be discussed first followed by the results for the 2004 Strong.

For the 1994 Strong, the overwhelming majority of tests (41 / 50 = 82%) garnered significant results ($p < .0005$) indicating a violation of the normality assumption. Eight of the nine 1994 scales which proved not to be statistically significant ($p > .0005$), and thus which can be considered to be normally distributed, occurred only for males. The following 1994 male BISs failed to show statistical significance ($p > .0005$): Science, Music/Dramatics, Social Service, Medical Service, Public Speaking, Merchandising, Organizational Management, and Data Management. The only 1994 BIS failing to find significance for the females was Merchandising.

Whereas the majority of the results for the 1994 BISs indicated non-normal distribution, the 2004 BISs showed almost the opposite results. That is, the majority of the results for the 2004 BISs, or 45 of the 60 tests (75%), supported normal distribution. Of the 15 statistically significant scales ($p < .0005$), 14 were significant only for females. The 14 statistically significant 2004 Strong female BISs included the following: Mechanics and Construction, Computer Hardware and Electronics, Military, Athletics, Science, Mathematics, Culinary Arts, Teaching and Education, Religion and Spirituality, Entrepreneurship, Politics and Public Speaking, Law, Taxes and Accounting, and Programming and Information Systems. The final significant 2004 BIS was the male Athletics scale.

Table 3.5

Shapiro-Wilks's Tests of Normality for the Basic Interest Scales of the 1994 and 2004 Strong Interest Inventories by Sex

		Shapiro-Wilks Tests		
BIS	**Sex**	**Statistic (W)**	**df**	***p*-value**
1994 Agriculture	Male	.970	248	*<.001*
	Female	.947	374	*<.001*
2004 Nature & Agriculture	Male	.992	248	.188
	Female	.991	374	.022
1994 Nature	Male	.976	248	*<.001*
	Female	.967	374	*<.001*
2004 Nature & Agriculture	Male	.992	248	.188
	Female	.991	374	.022
1994 Military Activities	Male	.925	248	*<.001*
	Female	.751	374	*<.001*
2004 Military	Male	.987	248	.025
	Female	.946	374	*<.001*
1994 Athletics	Male	.888	248	*<.001*
	Female	.954	374	*<.001*
2004 Athletics	Male	.947	248	*<.001*
	Female	.983	374	*<.001*
1994 Mechanical Activities	Male	.974	248	*<.001*
	Female	.836	374	*<.001*
2004 Mechanics & Construction	Male	.989	248	.061
	Female	.973	374	*<.001*
1994 Science	Male	.983	248	.005
	Female	.934	374	*<.001*
2004 Science	Male	.993	248	.323
	Female	.983	374	*<.001*
1994 Mathematics	Male	.966	248	*<.001*
	Female	.944	374	*<.001*
2004 Mathematics	Male	.984	248	.007
	Female	.976	374	*<.001*
1994 Medical Science	Male	.967	248	*<.001*
	Female	.957	374	*<.001*
2004 Medical Science	Male	.991	248	.139
	Female	.986	374	.001
1994 Music/Dramatics	Male	.983	248	.004
	Female	.982	374	*<.001*

Table 3.5 (Continued)

		Shapiro-Wilks Tests		
BIS	**Sex**	**Statistic (W)**	**df**	***p*-value**
2004 Performing Arts	Male	.991	248	.122
	Female	.987	374	.003
1994 Art	Male	.967	248	*<.001*
	Female	.970	374	*<.001*
2004 Visual Arts & Design	Male	.991	248	.142
	Female	.989	374	.005
1994 Applied Art	Male	.971	248	*<.001*
	Female	.972	374	*<.001*
2004 Performing Arts	Male	.991	248	.122
	Female	.987	374	.003
1994 Writing	Male	.952	248	*<.001*
	Female	.971	374	*<.001*
2004 Writing & Mass Comm.	Male	.988	248	.031
	Female	.988	374	.005
1994 Culinary Arts	Male	.966	248	*<.001*
	Female	.898	374	*<.001*
2004 Culinary Arts	Male	.989	248	.067
	Female	.972	374	*<.001*
1994 Teaching	Male	.971	248	*<.001*
	Female	.947	374	*<.001*
2004 Teaching & Education	Male	.993	248	.303
	Female	.981	374	*<.001*
1994 Social Service	Male	.987	248	.025
	Female	.971	374	*<.001*
2004 Counseling & Helping	Male	.986	248	.014
	Female	.993	374	.071
1994 Medical Service	Male	.985	248	.012
	Female	.972	374	*<.001*
2004 Healthcare Services	Male	.984	248	.006
	Female	.992	374	.038
1994 Religious Activities	Male	.930	248	*<.001*
	Female	.940	374	*<.001*
2004 Religion & Spirituality	Male	.978	248	.001
	Female	.980	374	*<.001*
1994 Public Speaking	Male	.980	248	.001
	Female	.965	374	*<.001*
2004 Politics & Public Speaking	Male	.991	248	.157
	Female	.983	374	*<.001*

Table 3.5 (Continued)

		Shapiro-Wilks Tests		
BIS	**Sex**	**Statistic (W)**	**df**	***p*-value**
1994 Law/Politics	Male	.977	248	*<.001*
	Female	.955	374	*<.001*
2004 Politics & Public Speaking	Male	.991	248	.157
	Female	.983	374	*<.001*
2004 Law	Male	.986	248	.016
	Female	.971	374	*<.001*
1994 Merchandising	Male	.983	248	.005
	Female	.989	374	.006
2004 Marketing & Advertising	Male	.995	248	.685
	Female	.989	374	.004
1994 Sales	Male	.965	248	*<.001*
	Female	.974	374	*<.001*
2004 Sales	Male	.990	248	.074
	Female	.986	374	.001
1994 Organizational Management	Male	.987	248	.028
	Female	.984	374	*<.001*
2004 Human Resources & Training	Male	.987	248	.026
	Female	.991	374	.021
1994 Data Management	Male	.982	248	.003
	Female	.941	374	*<.001*
2004 Management	Male	.990	248	.100
	Female	.992	374	.042
1994 Computer Activities	Male	.963	248	*<.001*
	Female	.886	374	*<.001*
2004 Programming & Info Systems	Male	.991	248	.125
	Female	.982	374	*<.001*
2004 Comp. Hardware & Electronics	Male	.983	248	.004
	Female	.940	374	*<.001*
1994 Office Services	Male	.966	248	*<.001*
	Female	.970	374	*<.001*
2004 Office Management	Male	.991	248	.110
	Female	.989	374	.006
2004 Protective Services	Male	.982	248	.004
	Female	.991	374	.027
2004 Research	Male	.994	248	.392
	Female	.990	374	.015
2004 Social Sciences	Male	.982	248	.003
	Female	.991	374	.021

Table 3.5 (Continued)

		Shapiro-Wilks Tests		
BIS	**Sex**	**Statistic (W)**	**df**	***p*-value**
2004 Entrepreneurship	Male	.985	248	.009
	Female	.984	374	*<.001*
2004 Taxes & Accounting	Male	.987	248	.024
	Female	.977	374	*<.001*
2004 Finance & Investing	Male	.988	248	.038
	Female	.989	374	.007

Note. All p-values in italics denote statistical significance at the .0005 level

Personal Style Scales. Table 3.6 presents the normality tests for the 1994 and 2004 PSSs by sex. Not surprisingly, the results display distinct differences between the 1994 and 2004 Strongs, with a greater number of the 2004 scales failing to find a Bonferroni adjusted statistical significance (.05 / 18 = .003). The following four scales were statistically significant ($p < .003$) for the 1994 Strong indicating a non-normal distribution: the female Work Style scale, the female Leadership Style scale, and both the female and male Risk Taking/Adventure scales. The remaining four BISs of the 1994 Strong were not statistically significant ($p > .003$) and included the following: the male Work Style scale, both the female and male Learning Environment scales, and the male Leadership Style scale. For the 2004 Strong, nine of the ten BISs failed to find statistical significance ($p > .003$), with the exception being the male Risk Taking scale ($p < .003$). That is, 90% of the results for the 2004 PSS normality tests indicated normal distribution.

Table 3.6

Shapiro-Wilks's Tests of Normality for the Personal Style Scales of the 1994 and 2004 Strong Interest Inventories by Sex

		Shapiro-Wilks Tests		
PSS	**Sex**	**Statistic (W)**	**df**	***p*-value**
1994 Work Style	Male	.993	248	.253
	Female	.986	374	*.001*
2004 Work Style	Male	.994	248	.446
	Female	.995	374	.285
1994 Learning Environment	Male	.992	248	.162
	Female	.996	374	.360
2004 Learning Environment	Male	.988	248	.044
	Female	.994	374	.125
1994 Leadership Style	Male	.986	248	.016
	Female	.982	374	*<.001*
2004 Leadership	Male	.993	248	.256
	Female	.996	374	.538
1994 Risk Taking/Adventure	Male	.949	248	*<.001*
	Female	.981	374	*<.001*
2004 Risk Taking	Male	.975	248	*<.001*
	Female	.993	374	.090
2004 Team Orientation	Male	.989	248	.056
	Female	.989	374	.009

Note. All p-values in italics denote statistical significance at the .003 level

Univariate Analysis of Sex Differences of the 2004 Strong Interest Inventory

Several studies have determined that sex differences are inherent in the construct of vocational interests (Hansen, Collins, Swanson, & Fouad, 1993; Holland & Gottfredson, 1976; Jackson, 1977; Johansson, 1986; Zytowski, 1985). This is evidenced in the Strong Interest Inventory by the fact that the Occupational Scales (not presented in this study) are normed separately for females and males. Thus, this analysis was not an attempt to determine if sex differences existed, but rather to what extent they existed. To accomplish this, the final preliminary analysis conducted in this study were a series of One-way between-groups univariate analyses of variance (ANOVAs) to investigate sex differences between the 1994 and 2004 Strongs.

Results of the 2004 Univariate Analyses of Variances for the GOTs, BISs, and PSSs by Sex

Tests of between-subjects effects on each dependent variable were examined to determine whether sex exuded a significant effect for each content scale. Once again, the Bonferroni method was employed to control for elevated Type I errors. The results of these univariate analyses of variance (ANOVAs) for each of the content scales are discussed in turn followed by a summary of overall sex effects between the 1994 and 2004 Strongs.

General Occupational Themes. Table 3.7 below presents the results of the tests of between-subjects effects for each of the GOTs for both the 1994 and 2004 Strongs. After the Bonferroni adjustment, each univariate ANOVA was tested at the .004 level (.05 / 12 = .004). Unlike the tests of normality discussed earlier, the univariate tests of between-subjects effects of sex differences were similar for both versions of the Strong. Specifically, for both the 1994 and 2004 Strongs, the ANOVAs on the Realistic (F [1, 620] = 324.04, $p < .001$; F [1, 620] = 278.36, $p < .001$, for the 1994 and 2004 SII respectively), Investigative (F [1, 620] =

38060, $p < .001$; F [1, 620] = 26.28, $p < .001$, for the 1994 and 2004 SII respectively), Artistic (F [1, 620] = 49.28, $p < .001$; F [1, 620] = 33.42, $p < .001$, for the 1994 and 2004 SII respectively), and Social (F [1, 620] = 62.22, $p < .001$; F [1, 620] = 65.21, $p < 001$, for the 1994 and 2004 SII respectively) scales were statistically significant denoting the presence of sex effects.

The non-significant scales differed only slightly between the two versions. The 1994 Strong had two scales that failed to show statistical significance, the first was for the Enterprising scale (F [1, 620] = .10, p = .758) and the second was for the Conventional scale (F [1, 620] = 2.80, p = .095). For the 2004 Strong, only the Enterprising GOT failed to show significance (F [1, 620] = .51, p = .477). Again using Cohen's (1988) criteria, an examination of the partial eta-squared values revealed strong effect sizes for both the 1994 (η_p^2 = .34) and the 2004 (η_p^2 = .31) Realistic GOT. Medium effect sizes were observed for the 1994 Investigative (η_p^2 = .06), Artistic (η_p^2 = .07), and Social (η_p^2 = .09) GOTs; as well as the 2004 Social GOT (η_p^2 = .10). Finally, small effect sizes were observed for the 2004 Investigative (η_p^2 = .04), Artistic (η_p^2 = .05), and Conventional (η_p^2 = .04) GOTs.

Table 3.7

Tests of Between-Subjects Effects of Sex Differences between the 1994 and 2004 Strong Interest Inventories General Occupational Themes

Source	Dependent Variable (GOT)	F	Sig.	η_p^2	Dependent Variable (GOT)	F	Sig.	η_p^2
Sex	1994 Realistic	324.04	*.000*	.343	2004 Realistic	278.36	*<.001*	.310
	1994 Investigative	38.60	*.000*	.059	2004 Investigative	26.28	*<.001*	.041

Table 3.7 (Continued)

Source	Dependent Variable (GOT)	F	Sig.	η_p^2	Dependent Variable (GOT)	F	Sig.	η_p^2
Sex	1994 Artistic	49.28	*.000*	.074	2004 Artistic	33.42	*<.001*	.051
	1994 Social	62.22	*.000*	.091	2004 Social	65.21	*<.001*	.095
	1994 Enterprising	.10	.758	.000	2004 Enterprising	.51	.477	.001
	1994 Conventional	2.80	.095	.004	2004 Conventional	24.70	*<.001*	.038

Note. Significance values in italics denote statistical significance at the $p < .004$ level

Basic Interest Scales. Table 3.8 delineates the results of the tests of between-subjects sex effects for each of the GOTs for both the 1994 and 2004 Strongs. Parallel to the GOTs, after the Bonferroni adjustment, each univariate BIS ANOVA was tested at the .001 level (.05 / 55 = .001). Also as with the GOTs, the results for the 1994 and 2004 Strongs were remarkably similar, as expected. Overall, 17 of the 25 (or 68%) of the ANOVAs for the 1994 BISs were statistically significant ranging from Medical Service (F [1, 620] = .11.39, $p <$ .001) to Mechanical Activities (F [1, 620] = 350.38, $p < .001$). The 2004 Strong showed a similar proportion of results with 22 out of 30 (or 73%) of its BISs being statistically significant ranging from Entrepreneurship (F [1, 620] = 12.32, $p < .001$) to Mechanics and Construction (F [1, 620] = 294.99, p < .001). Thus, the pattern of sex effects was similar across both versions of the Strong.

The pattern of non-significant results also showed striking similarities between the 1994 and 2004 BISs. Six of the nine non-significant 1994 BIS ANOVAs ($p > .001$) were analogous to non-significant 2004 BIS ANOVAs ($p > .001$). These twelve non-significant BISs included the following: 1) 1994 Medical Science with 2004 Medical Science, 2) 1994 Applied Art with 2004 Visual Arts and Design, 3) 1994 Religious Activities with 2004

Religion and Spirituality, 4) 1994 Merchandising with 2004 Marketing and Advertising, 5) 1994 Organizational Management with 2004 Office Management, and 6) 1994 Office Services with 2004 Office Management. The three 1994 non-significant BIS scales were Nature (F [1, 620] = .48, p = .49), Teaching (F [1, 620] = 6.03, p = .01), and Sales (F [1, 620] = 4.54, p < .03). The two non-analogous 2004 non-statistically significant BISs were Social Services (F [1, 620] = 1.95, p = .16) and Law (F [1, 620] = 7.69, p = .01).

Of note, however are three sets of 1994 and 2004 BISs which demonstrated opposing results. The 1994 Applied Arts BIS unmistakably failed to show significance (p = .813), while the 2004 Visual Arts and Design BIS just missed showing significance but (p = .002). Next, the 1994 Teaching BIS showed no significant mean differences between sexes (p = .014), while the parallel 2004 Teaching and Education BIS did show a significant mean difference (p < .001). Finally, the 1994 Sales BIS did not display significance (p = .034), while its analogous 2004 Sales BIS did demonstrate a statistically significant mean difference between sexes (p < .001).

Turning once more to Cohen's (1988) criteria, an examination of the partial eta-squared values revealed large effect sizes ($\eta_p^2 \geq .14$) for three of the 1994 Strong and for four of the 2004 Strong BISs. Medium effect sizes ($.14 > \eta_p^2 \geq .06$) were observed for nine of the 1994 and for eight of the 2004 Strong BISs. Finally, small effect sizes ($\eta_p^2 \leq .04$) were observed for four of the 1994 BISs and for nine of the 2004 BISs. See Table 3.8 below for a complete delineation of the tests of between-subjects effects of sex differences between the 1994 and 2004 Strong BISs.

Table 3.8

Tests of Between-Subjects Effects of Sex Differences between the 1994 and 2004 Strong Interest Inventories Basic Interest Scales

Source	Dependent Variable (BIS)	F	Sig.	η_p^2
Sex	1994 Agriculture	34.46	<.001	.053
	2004 Nature & Agriculture	12.73	<.001	.020
	1994 Nature	.48	.487	.001
	2004 Nature & Agriculture	12.73	<.001	.020
	1994 Military Activities	143.60	<.001	.188
	2004 Military	151.24	<.001	.196
	1994 Athletics	45.55	<.001	.068
	2004 Athletics	23.95	<.001	.037
	1994 Mechanical Activities	350.38	<.001	.361
	2004 Mechanics & Construction	294.99	<.001	.322
	1994 Science	60.79	<.001	.089
	2004 Science	41.26	<.001	.062
	1994 Mathematics	85.65	<.001	.121
	2004 Mathematics	59.83	<.001	.088
	1994 Medical Science	1.71	.190	.003
	2004 Medical Science	2.11	.147	.003
	1994 Music/Dramatics	37.19	<.001	.057
	2004 Performing Arts	52.79	<.001	.078
	1994 Art	42.69	<.001	.064
	2004 Visual Arts & Design	9.25	.002	.015
	1994 Applied Arts	.06	.813	.000
	2004 Visual Arts & Design	9.25	.002	.015
	1994 Writing	26.51	<.001	.041
	2004 Writing & Mass Communication	15.00	<.001	.024
	1994 Culinary Arts	90.42	<.001	.127
	2004 Culinary Arts	36.41	<.001	.055
	1994 Teaching	6.03	.014	.010
	2004 Teaching & Education	26.42	<.001	.041
	1994 Social Service	122.35	<.001	.165
	2004 Counseling & Helping	70.77	<.001	.102
	1994 Medical Service	11.39	.001	.018
	2004 Healthcare Services	13.16	<.001	.021
	1994 Religious Activities	3.01	.083	.005
	2004 Religion & Spirituality	1.95	.163	.003

Table 3.8 (Continued)

Source	Dependent Variable (BIS)	F	Sig.	η_p^2
Sex	1994 Public Speaking	12.34	*<.001*	.020
	2004 Politics & Public Speaking	25.71	*<.001*	.040
	1994 Law/Politics	59.55	*<.001*	.088
	2004 Politics & Public Speaking	25.71	*<.001*	.040
	2004 Law	7.69	.006	.012
	1994 Merchandising	3.55	.060	.006
	2004 Marketing & Advertising	7.42	.007	.012
	1994 Sales	4.54	.034	.007
	2004 Sales	16.05	*<.001*	.025
	1994 Organizational Management	9.92	.002	.016
	2004 Human Resources & Training	5.10	.024	.008
	1994 Data Management	45.66	*<.001*	.069
	2004 Management	24.38	*<.001*	.038
	1994 Computer Activities	64.98	*<.001*	.095
	2004 Computer Hardware & Electronics	192.88	*<.001*	.237
	2004 Programming & Information Systems	69.85	*<.001*	.101
	1994 Office Services	5.31	.022	.008
	2004 Office Management	6.00	.015	.010
	2004 Protective Services	98.53	*<.001*	.137
	2004 Research	24.46	*<.001*	.038
	2004 Social Sciences	1.24	.267	.002
	2004 Entrepreneurship	12.32	*<.001*	.019
	2004 Taxes & Accounting	39.22	*<.001*	.059
	2004 Finance & Investing	66.59	*<.001*	.097

Note. Significance values in italics denote statistical significance at the $p < .001$ level

Personal Style Scales. Table 3.9 delineates the results of the tests of between-subjects sex effects for each of the GOTs for both the 1994 and 2004 Strongs. Parallel to the GOTs, after the Bonferroni adjustment, each univariate BIS ANOVA was tested at the .006 level (.05 / 9 = .006). The PSSs showed the most similarity between the 1994 and 2004 Strongs. More precisely, the univariate ANOVAs resulted in the Work Style and Risk Taking PSSs being statistically significant ($p < .006$), and in the Learning Environment and Leadership PSSs being statistically non-significant ($p > .006$) for both the 1994 and 2004 Strongs. The

additional Team Orientation scale on the 2004 Strong was non-significant (F [1, 620] = .861, p = .354). Effect sizes were also parallel with both the 1994 and 2004 Work Style and Risk Taking scales having large effect sizes (Cohen, 1988; $\eta_p^2 \geq .14$), and all remaining scales for both versions showing small effect sizes ($\eta_p^2 \leq .04$).

Table 3.9

Tests of Between-Subjects Effects of Sex Differences between the 1994 and 2004 Strong Interest Inventories Personal Style Scales

Source	Dependent Variable (PSS)	F	Sig.	η_p^2	Dependent Variable (PSS)	F	Sig.	η_p^2
Sex	1994 Work Style	271.75	*<.001*	.305	2004 Work Style	244.03	*<.001*	.282
	1994 Learning Environment	3.69	.055	.006	2004 Learning Environment	.06	.807	<.001
	1994 Leadership Style	1.39	.239	.002	2004 Leadership	.67	.415	.001
	1994 Risk Taking/Adventure	116.06	*<.001*	.158	2004 Risk Taking	133.94	*<.001*	.178
					2004 Team Orientation	.86	.354	.001

Note. Significance values in italics denote statistical significance at the p < .006 level

Summary of Results for the Tests of Between-Subjects Effects for Sex. The strength of the similarity between the results of the two versions of the Strong further buttresses the argument for the current sample being an adequate representation of the 2004 GRS. The close parallel between the 1994 and 2004 content scales that demonstrated sex effects and the comparable effect sizes for these statistically significant scales provides evidence of the equivalence of sex differences between the two versions. As with their significant

counterparts, the similarity in non-significant content scales between the 1994 and 2004 Strongs failing to find sex effects offers support for the first hypothesis.

Paired Sample t-tests between Analogous Content Scales of the 1994 and 2004 Strong Interest Inventories: Hypothesis One

Paired samples t-tests were used to test the first hypothesis that no meaningful differences would exist between the means of the parallel content scales of the 1994 and 2004 Strong Interest Inventories. Specifically, it was hypothesized that no mean differences greater than one-half standard deviation would exist between any analogous scales. The paired-sample t-tests were computed separately for each sex. Each set of content scales will be discussed in turn.

Paired Sample t-tests between General Occupational Themes of the 1994 and 2004 Strong by Sex

Female GOTs. Table 3.11 presents the paired sample t-tests between the GOTs of the 1994 and 2004 Strongs. As with all previous analyses, Bonferroni adjustments were made to minimize the likelihood of Type I errors resulting in the construction of 99.6% confidence intervals. The means and standard deviations of the 1994 GOTs for males and females are presented in Table 3.10. As can be seen from Table 3.11, females scored significantly higher on four of the GOT mean scores from the 1994 to the 2004 Strongs including the following: 1) Realistic, $t(373) = 24.85$, $p < .0005$, 2) Investigative, $t(373) = 26.79$, $p < .0005$, 3) Artistic, $t(373) = 21.23$, $p < .0005$, and 4) Social, $t(373) = 4.23$, $p < .0005$. The 2004 Enterprising GOT mean was significantly lower than the 1994 Enterprising GOT, $t(373) = -5.96$, $p < .0005$. The final GOT, Conventional, was not significant.

Male GOTs. Other than showing significance on the Conventional GOT, the male paired-samples t-tests do not appear to deviate strongly from the female pattern. Specifically, as seen in Table 3.11, five of the 2004 GOT means were significantly higher than the 1994 means for the following scales: 1) Realistic, $t(247) = 14.11$, $p < .0005$, 2) Investigative, $t(274) = 18.85$, $p < .0005$, 3) Artistic, $t(274) = 24.12$, $p < .0005$, 4) Social, $t(274) = 5.54$, $p < .0005$, and 5) Conventional, $t(274) = 10.48$, $p < .0005$. The result for the Enterprising scale was in the opposite direction with the 2004 Strong GOT mean being significantly lower than the 1994 Strong GOT mean, $t(274) = -3.24$, $p = .001$.

Table 3.10

Means and Standard Deviations of the General Occupational Themes of the 1994 and 2004 Strong Interest Inventories by Sex (females n =374; males n = 248)

Sex	General Occupational Theme	M	SD	General Occupational Theme	M	SD
F	1994 Realistic	41.86	8.17	2004 Realistic	47.52	8.01
M	1994 Realistic	54.66	9.41	2004 Realistic	58.44	7.98
F	1994 Investigative	43.53	9.45	2004 Investigative	48.24	9.51
M	1994 Investigative	48.33	9.44	2004 Investigative	52.18	9.18
F	1994 Artistic	49.54	9.32	2004 Artistic	52.57	8.83
M	1994 Artistic	44.16	9.42	2004 Artistic	48.37	8.92
F	1994 Social	55.67	10.43	2004 Social	56.58	9.59
M	1994 Social	49.13	9.65	2004 Social	50.49	8.61
F	1994 Enterprising	54.12	9.86	2004 Enterprising	52.80	9.64
M	1994 Enterprising	54.38	10.89	2004 Enterprising	53.36	9.50
F	1994 Conventional	50.79	10.47	2004 Conventional	51.44	10.82
M	1994 Conventional	52.20	10.02	2004 Conventional	55.74	10.34

Summary of GOT Paired Sample t-tests. Though all the paired sample t-tests were statistically significant except for the female Conventional GOT means, a closer examination of the mean differences reveals partial support for the first hypothesis. The female Realistic GOT showed the greatest magnitude in the difference between the means of the 1994 and 2004 Strong (Mean difference = 5.66), which just falls at the five-point, half a standard deviation hypothesis cut-off. All other statistically significant GOTs resulted in differences in their means that were less than five points.

Table 3.11

Paired Sample t-tests between the General Occupational Themes of the 1994 and 2004 Strong Interest Inventories by Sex

	Paired Differences						
			99.6% CI				
Version (by Year)	**Mean Difference**	**SD**	**Lower**	**Upper**	***t***	**df**	**Sig.**
2004-1994 Realistic (F)	5.66	4.40	5.00	6.32	24.85	373	*<.001*
2004-1994 Realistic (M)	3.78	4.22	3.00	4.56	14.11	247	*<.001*
2004-1994 Investigative (F)	4.72	3.40	4.21	5.23	26.79	373	*<.001*
2004-1994 Investigative (M)	3.85	3.21	3.25	4.44	18.85	247	*<.001*
2004-1994 Artistic (F)	3.03	2.76	2.61	3.44	21.23	373	*<.001*
2004-1994 Artistic (M)	4.21	2.75	3.70	4.72	24.12	247	*<.001*
2004-1994 Social (F)	.91	4.16	.29	1.53	4.23	373	*<.001*
2004-1994 Social (M)	1.35	3.87	.65	2.07	5.54	247	*<.001*
2004-1994 Enterprising (F)	1.31	4.26	-1.95	-.67	-5.96	373	*<.001*
2004-1994 Enterprising (M)	-1.01	4.92	-1.92	-.11	-3.24	247	*< .001*
2004-1994 Conventional (F)	.63	5.03	-.13	1.38	2.40	373	.017
2004-1994 Conventional (M)	3.54	5.32	2.56	4.52	10.48	247	*<.001*

Note. Significance values in italics denote statistical significance at the $p < .004$ level

Paired Sample t-tests between Analogous Basic Interest Scales of the 1994 and 2004 Strong by Sex

Female BISs. Table 3.13 presents the paired sample t-tests between the BISs of the 1994 and 2004 Strongs. Again, Bonferroni adjustments were made to decrease the occurrence of Type I errors resulting in the construction of 99.9% confidence intervals for the BISs. Means and standard deviations for both females and males are presented in Table 3.12. For females, there was a statistically significant difference in 19 of the 22 BIS mean scores from the 1994 to the 2004 Strongs. Of those differences, 16 of the 19 differences revealed the 2004 BIS means were significantly higher than the 1994 BIS means. Thus, proportionally fewer ($n = 3$) female BIS paired-sample t-tests resulted in the 2004 BIS mean differences being significantly lower than the 1994 BIS means. These included the following scales: 1) Culinary Arts, $t(373) = -10.08$, $p < .0005$, 2) 1994 Social Service versus 2004 Counseling and Helping, $t(373) = -10.04$, $p < .0005$, and 3) Sales, $t(373) = -8.52$, $p < .0005$. Finally, three BIS mean differences (i.e., Military, Religion and Spirituality, and Office Management) failed to show significance at $p < .001$ level.

Male BISs. As shown in Table 3.13, 14 of the 22 BIS means were statistically different for males. The 13 BIS means from the 2004 Strong were significantly higher than the 1994 BIS means as shown by tables 3.12 and 3.13. Only the Military BIS mean was significantly lower on the 2004 Strong compared to the 1994 Strong.

Table 3.12

Means and Standard Deviations of the Basic Interest Scales of the 1994 and 2004 Strong Interest Inventories by Sex (females n =374; males n = 248)

Sex	Basic Interest Scale	M	SD	Basic Interest Scale	M	SD
F	1994 Mechanical Activ	42.60	7.71	2004 Mechanics/Const	46.16	7.74
M	1994 Mechanical Activ	55.71	9.69	2004 Mechanics/Const	57.76	8.95
F	1994 Military Activities	48.37	8.60	2004 Military	47.76	8.23
M	1994 Military Activities	55.71	9.69	2004 Military	56.63	9.61
F	1994 Nature	44.32	10.19	2004 Nature/Agriculture	49.92	9.12
M	1994 Nature	44.88	9.18	2004 Nature/Agriculture	52.48	8.19
F	1994 Agriculture	43.69	9.08	2004 Nature/Agriculture	49.92	9.12
M	1994 Agriculture	48.08	9.16	2004 Nature/Agriculture	52.48	8.19
F	1994 Athletics	53.12	8.54	2004 Athletics	55.56	9.08
M	1994 Athletics	57.73	8.03	2004 Athletics	59.20	9.11
F	1994 Science	43.63	8.60	2004 Science	47.97	9.26
M	1994 Science	49.23	9.00	2004 Science	52.77	8.91
F	1994 Medical Science	49.50	10.72	2004 Medical Science	53.68	10.63
M	1994 Medical Science	50.62	9.98	2004 Medical Science	54.88	9.32
F	1994 Mathematics	44.31	8.95	2004 Mathematics	49.49	9.26
M	1994 Mathematics	51.24	9.44	2004 Mathematics	55.33	9.15
F	1994 Art	50.24	9.51	2004 Visual Arts/Design	51.77	9.11
M	1994 Art	45.08	9.85	2004 Visual Arts/Design	49.50	9.13
F	1994 Music/Dramatics	51.92	9.34	2004 Performing Arts	53.53	9.34
M	1994 Music/Dramatics	47.31	9.09	2004 Performing Arts	47.82	9.99
F	1994 Writing	46.22	9.75	2004 Writing/Mass Comm	49.72	9.92
M	1994 Writing	42.18	9.32	2004 Writing/Mass Comm	46.67	9.10
F	1994 Culinary Arts	55.52	8.49	2004 Culinary Arts	52.95	8.44
M	1994 Culinary Arts	48.67	9.24	2004 Culinary Arts	48.71	8.82
F	1994 Social Service	57.20	9.58	2004 Counseling/Helping	53.97	8.93
M	1994 Social Service	48.71	9.04	2004 Counseling/Helping	47.99	8.28
F	1994 Teaching	50.13	10.45	2004 Teaching/Education	56.87	10.64
M	1994 Teaching	48.06	10.02	2004 Teaching/Education	52.59	9.38

Table 3.12 (Continued)

Sex	Basic Interest Scale	M	SD	Basic Interest Scale	M	SD
F	1994 Religious Activities	50.79	10.26	2004 Religion/Spirituality	50.81	9.61
M	1994 Religious Activities	49.33	10.32	2004 Religion/Spirituality	49.72	9.45
F	1994 Medical Service	55.19	11.49	2004 Healthcare Services	57.09	10.49
M	1994 Medical Service	52.17	9.97	2004 Healthcare Services	54.16	8.84
F	1994 Sales	56.52	10.56	2004 Sales	54.00	10.68
M	1994 Sales	58.47	12.09	2004 Sales	57.52	10.84
F	1994 Data Management	46.94	9.05	2004 Management	49.88	9.63
M	1994 Data Management	52.00	9.26	2004 Management	53.76	9.54
F	1994 Law/Politics	45.69	10.07	2004 Politics/Pub Speaking	48.77	9.75
M	1994 Law/Politics	51.91	9.48	2004 Politics/Pub Speaking	52.67	8.79
F	1994 Public Speaking	47.63	9.96	2004 Politics/Pub Speaking	48.77	9.75
M	1994 Public Speaking	50.44	9.48	2004 Politics/Pub Speaking	52.67	8.79
F	1994 Law/Politics	45.69	10.07	2004 Law	50.64	10.16
M	1994 Law/Politics	51.91	9.48	2004 Law	52.87	9.21
F	1994 Office Services	54.04	10.37	2004 Office Management	54.52	10.52
M	1994 Office Services	52.16	9.36	2004 Office Management	52.51	9.21

Summary of BIS Paired Sample t-tests. Seven of the 22 analogous BIS paired sample t-tests garnered opposite statistical significance for females and males. Of these seven, six were statistically significant ($p < .001$) for females but not males (i.e., Performing Arts, Culinary Arts, Counseling & Helping, Sales, Politics & Public Speaking, and Law). Only the Military BIS pair proved to be statistically significant for males but not for females. In total, 44 paired-samples t-tests were conducted to determine the existence of mean differences in scores on the analogous BIS between the 1994 and 2004 Strong. It was hypothesized that no

statistically significant mean differences would be greater than one-half (or five points) of a standard deviation. Of the 33 statistically significant mean differences ($p < .001$), seven were equal to or greater than one-half a standard deviation (i.e., mean difference > 4.95), with the scale with the greatest differential magnitude (the 2004/1994 Nature and Agriculture with Nature male BISs) displaying a difference of more than one-half of a standard deviation (mean difference = 7.60).

Table 3.13

Paired Sample t-tests between the Analogous Basic Interest Scales of the 1994 and 2004 Strong Interest Inventories by Sex

	Paired Differences						
			99.9% CI				
Version (by Year)*	**Mean Difference**	**SD**	**Lower**	**Upper**	***t***	**df**	**Sig.**
2004-1994 Mechanics/Const (F)	3.57	3.81	2.91	4.23	18.11	373	*<.001*
2004-1994 Mechanics/Const (M)	2.04	3.68	1.26	2.82	8.73	247	*<.001*
2004-1994 Military (F)	-.60	4.59	-1.39	.18	-2.54	373	.011
2004-1994 Military (M)	-1.54	4.75	-2.54	-.54	-5.10	247	*<.001*
2004-1994 Nature & Ag. w/ Nature (F)	**5.60**	5.18	4.71	6.49	20.89	373	*<.001*
2004-1994 Nature & Ag. w/ Nature (M)	**7.60**	4.92	6.56	8.64	24.30	247	*<.001*
2004-1994 Nature & Ag. w/ Ag (F)	**6.22**	7.31	4.97	7.48	16.46	373	*<.001*
2004-1994 Nature & Ag. w/ Ag (M)	4.40	7.47	2.82	5.98	9.28	247	*<.001*
2004-1994 Athletics (F)	2.44	3.39	1.86	3.02	13.94	373	*<.001*
2004-1994 Athletics (M)	1.47	3.47	.74	2.21	6.69	247	*<.001*
2004-1994 Science (F)	4.33	3.96	3.65	5.01	21.14	373	*<.001*
2004-1994 Science (M)	3.54	3.99	2.70	4.38	13.98	247	*<.001*
2004-1994 Medical Science (F)	4.18	5.35	3.26	5.10	15.09	373	*<.001*
2004-1994 Medical Science (M)	4.26	4.84	3.23	5.28	13.85	247	*<.001*
2004-1994 Mathematics (F)	**5.12**	3.61	4.57	5.80	27.79	373	*<.001*
2004-1994 Mathematics (M)	4.09	3.43	3.36	4.81	18.78	247	*<.001*

Table 3.13 (Continued)

Version (by Year)*	Mean Difference	SD	Lower	Upper	*t*	df	Sig.
2004-1994 Visual Art & Design (F)	**5.68**	4.74	4.87	6.50	23.16	373	*<.001*
2004-1994 Visual Art & Design (M)	3.60	4.25	2.70	4.50	13.36	247	*<.001*
2004-1994 Performing Arts (F)	1.61	3.88	.94	2.28	8.03	373	*<.001*
2004-1994 Performing Arts (M)	.51	4.16	-.37	1.39	1.91	247	.057
2004-1994 Writing & Mass Comm (F)	3.49	4.35	2.75	4.24	15.52	373	*<.001*
2004-1994 Writing & Mass Comm (M)	4.49	4.55	3.53	5.45	15.52	247	*<.001*
2004-1994 Culinary Arts (F)	-2.56	4.92	-3.40	-1.72	-10.08	373	*<.001*
2004-1994 Culinary Arts (M)	.04	4.47	-.91	.99	.14	247	.885
2004-1994 Counseling & Helping (F)	-3.22	6.21	-4.29	-2.16	-10.04	373	*<.001*
2004-1994 Counseling & Helping (M)	-.72	5.83	-1.95	.51	-1.94	247	.053
2004-1994 Teaching & Education (F)	**6.74**	4.49	5.97	7.51	29.01	373	*<.001*
2004-1994 Teaching & Education (M)	4.53	4.87	3.50	5.57	14.65	247	*<.001*
2004-1994 Religion & Spirituality (F)	.02	3.89	-.64	.69	.11	373	.912
2004-1994 Religion & Spirituality (M)	.39	3.94	-.44	1.23	1.57	247	.119
2004-1994 Healthcare Services (F)	1.91	4.22	1.18	2.63	8.74	373	*<.001*
2004-1994 Healthcare Services (M)	1.99	4.29	1.09	2.90	7.32	247	*<.001*
2004-1994 Sales (F)	-2.52	5.72	-3.50	-1.54	-8.52	373	*<.001*
2004-1994 Sales (M)	-.95	5.89	-2.19	.30	-2.54	247	.012
2004-1994 Management (F)	1.11	5.47	.17	2.05	3.93	373	*<.001*
2004-1994 Management (M)	2.39	5.55	1.22	3.57	6.79	247	*<.001*
2004-1994 Pol/Pub Speak w/ Law/Pol (F)	3.08	5.20	2.19	3.98	11.46	373	*<.001*
2004-1994 Pol/Pub Speak w/ Law/Pol (M)	.76	5.49	-.40	1.92	2.17	247	.031
2004-1994 Pol/Pub Speak w/ Pub Speak (F)	1.15	4.60	.36	1.93	4.81	373	*<.001*
2004-1994 Pol/Pub Speak w/ Pub Speak (M)	2.23	4.70	1.24	3.22	7.47	247	*<.001*
2004-1994 Law (F)	**4.96**	6.60	3.82	6.09	14.52	373	*<.001*
2004-1994 Law (M)	.96	5.99	-.31	2.23	2.53	247	.012
2004-1994 Office Management (F)	.48	5.34	.44	1.40	1.74	373	.083
2004-1994 Office Management (M)	.35	5.35	-.78	1.48	1.04	247	.301

*Note. * Except where noted, only 2004 scale name is listed, however all comparisons*

comprise 2004 and 1994 analogous BIS pairs presented in Table 2.4. Significance values in italics denote statistical significance at the p < .001 level. Mean differences in bold denote mean differences equal to or greater than one-half standard deviation.

Paired Sample t-tests between Analogous Personal Style Scales of the 1994 and 2004 Strong by Sex

Female PSSs. Table 3.15 presents the paired mean differences between the four analogous PSSs of the 1994 and 2004 Strongs. The means and standard deviations for both females and males are presented in Table 3.14. There were three statistically significant mean differences in PSS pairs for both females and males, and one non-significant mean difference per pair. Specifically, for females two 2004 PSS means were significantly higher than the 1994 PSS means. Namely the Learning Environment scale and the 1994 Risk Taking/Adventure scales versus the 2004 Risk Taking scale. The 2004 Leadership PSS was significantly lower than the 1994 Leadership PSS. The final female PSS, Work Style, did not display a statistically significant mean difference between Strong versions.

Male PSSs. As with the females, the male paired sample t-tests also resulted in two PSS showing a significantly difference in mean scores between the 1994 and 2004 Strongs. Similar to the females, the two 2004 PSS means that were higher than the 1994 means were the Learning Environment scale and the 1994 Risk Taking/Adventure scales versus the 2004 Risk Taking scale. However, both remaining PSS means (i.e., Work Style and Leadership) were not significantly different ($p > .006$).

Table 3.14

Means and Standard Deviations of the Personal Style Scales of the 1994 and 2004 Strong Interest Inventories by Sex (female n =374; male n = 248)

Sex	Personal Style Scale	M	SD	Personal Style Scales	M	SD
F	1994 Work Style	58.85	7.40	2004 Work Style	59.08	8.77
M	1994 Work Style	48.71	7.69	2004 Work Style	48.18	8.12
F	1994 Learning Environ	39.94	9.31	2004 Learning Environ	44.65	8.42
M	1994 Learning Environ	41.30	7.48	2004 Learning Environ	44.48	7.40
F	1994 Leadership Style	49.90	9.64	2004 Leadership	48.57	9.80
M	1994 Leadership Style	48.98	9.39	2004 Leadership	49.21	8.98
F	1994 Risk Taking/Advent	49.05	9.48	2004 Risk Taking	51.17	8.52
M	1994 Risk Taking/Advent	57.04	8.40	2004 Risk Taking	59.28	8.61

Summary of PSS Paired Samples t-tests. In total, eight paired sample t-tests were run on the PSSs by sex. Five of the eight t-tests resulted in a statistically significant difference between mean scores on the two versions of the Strong, three female PSSs and two male PSSs. All significant mean differences fell below one-half a standard deviation as expected from the first hypothesis.

Table 3.15

Paired Sample t-tests between the Analogous Personal Style Scales of the 1994 and 2004 Strong Interest Inventories

	Paired Differences						
			99.375% CI				
Version (by Year)	**Mean Differences**	**SD**	**Lower**	**Upper**	***t***	**df**	**Sig.**
2004-1994 Work Style (F)	.23	3.57	-.28	.74	1.24	373	.215
2004-1994 Work Style (M)	-.52	2.87	-1.03	-.02	-2.87	247	.004

Table 3.15 (Continued)

Version (by Year)	Mean Differences	SD	Lower	Upper	*t*	df	Sig.
2004-1994 Learning Environment (F)	4.70	4.49	4.07	5.34	20.28	373	*<.001*
2004-1994 Learning Environment (M)	3.19	4.80	2.35	4.03	10.45	247	*<.001*
2004-1994 Leadership (F)	-1.32	3.14	-1.77	-.87	-8.13	373	*<.001*
2004-1994 Leadership (M)	.23	3.09	-.31	.77	1.18	247	.239
2004-1994 Risk Taking (F)	2.12	6.49	1.20	3.05	6.32	373	*<.001*
2004-1994 Risk Taking (M)	2.24	6.47	1.10	3.37	5.44	247	*<.001*

Note. Significance values in italics denote statistical significance at the p < .006 level

Bivariate Correlations between Analogous Content Scales of the 1994 and 2004 Strong Interest Inventories: Hypothesis Two

Bivariate Correlations between General Occupational Scales of the 1994 and 2004 Strong Interest Inventories by Sex

Table 3.16 presents the correlations between the 1994 SII GOTs and the 2004 SII GOTs by sex. Interestingly, there appears to be no differentiation by sex as male correlations and female correlations are nearly identical for parallel scales. As expected, the highest correlations fell between parallel scales. Specifically, the highest correlation for both sexes was along the Artistic theme (r = .96). The lowest correlation between GOTs differed by sex with the females' lowest correlation falling on the Realistic theme (r = .85) and the males' lowest correlation falling on the Conventional theme (r = .86). However, due to the restricted range in overall correlates, this difference does not appear to be meaningful. The results of these correlations between the 1994 and 2004 Strongs support the second hypothesis and

provide evidence of no meaningful differences existing between parallel GOTs. Specifically, all parallel GOT correlations were above .85 as expected. Figures 3 through 7 below present the scatterplots of the parallel GOT scales, providing a graphical representation of the strong positive correlations between the 1994 and 2004 Strong GOTs.

Table 3.16

Correlations *between General Occupational Themes of the 1994 and 2004 Strong Interest Inventories by Sex (n = 622)*

	Females (n = 374)						Males (n = 248)					
	2004 GOTs											
1994 GOTs	R	I	A	S	E	C	R	I	A	S	E	C
R	**85**	50	43	27	33	37	**90**	48	12	23	27	20
I	44	**94**	27	15	10	33	45	**94**	33	36	10	30
A	35	24	**96**	27	33	06	18	33	**96**	52	28	16
S	13	23	25	**92**	49	31	23	42	51	**92**	49	45
E	14	04	35	37	**91**	44	22	15	23	41	**89**	53
C	34	27	10	21	66	**89**	24	32	10	33	66	**86**

Note. Decimals omitted. All parallel bold correlations are significant at the p < .01 level. Abbreviations for scales: R=Realistic, I=Investigative, A=Artistic, S=Social, E=Enterprising, C=Conventional.

Figure 3. Scatterplot of 2004 and 1994 Realistic General Occupational Themes

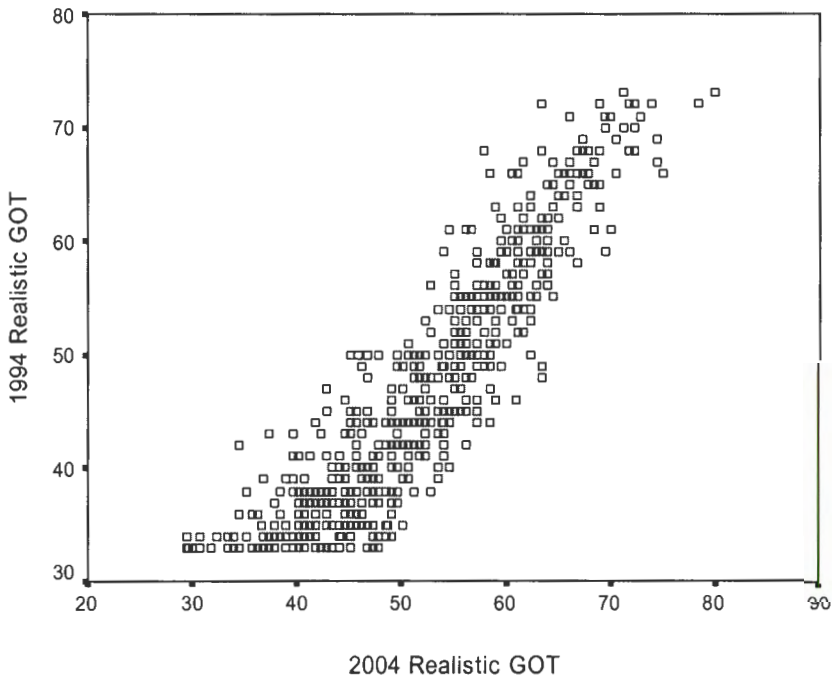


Figure 4. Scatterplot of 2004 and 1994 Investigative General Occupational Themes

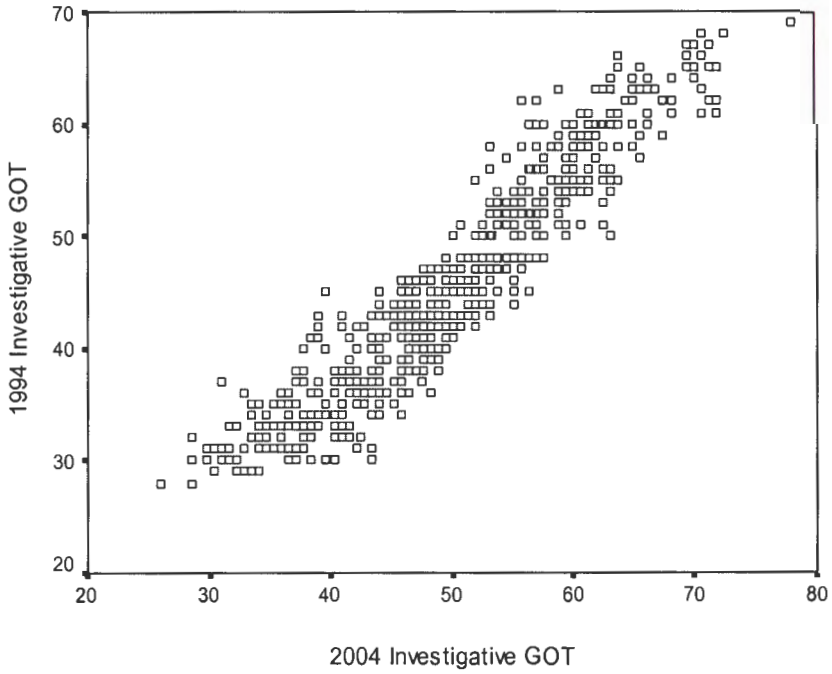


Figure 5. Scatterplot of 2004 and 1994 Artistic General Occupational Themes

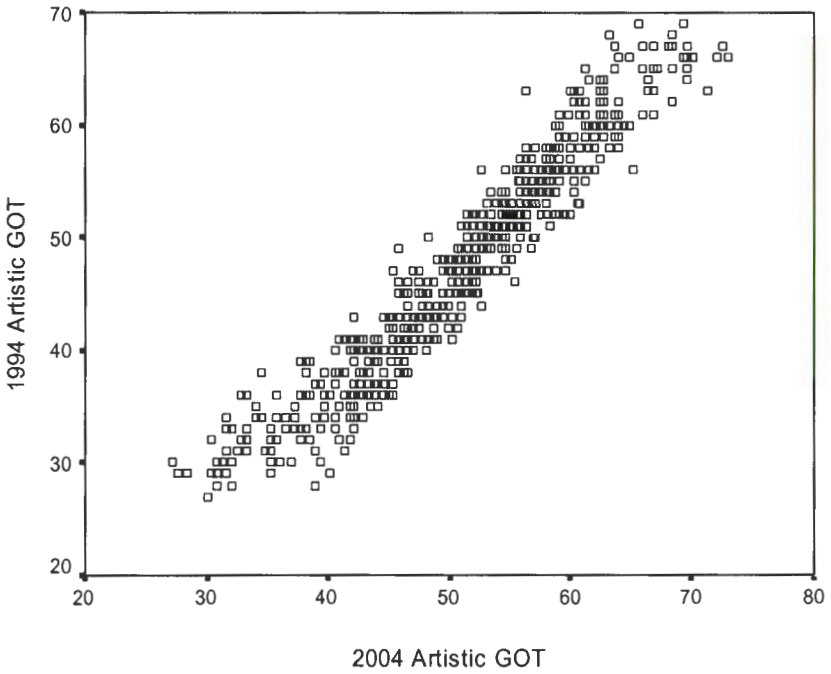


Figure 6. Scatterplot of 2004 and 1994 Social General Occupational Themes

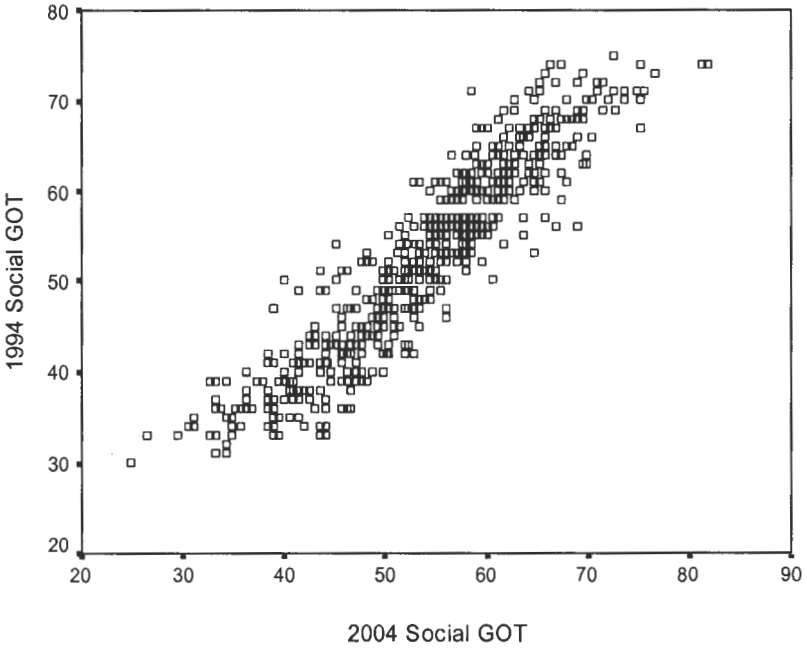


Figure 7. Scatterplot of 2004 and 1994 Enterprising General Occupational Themes

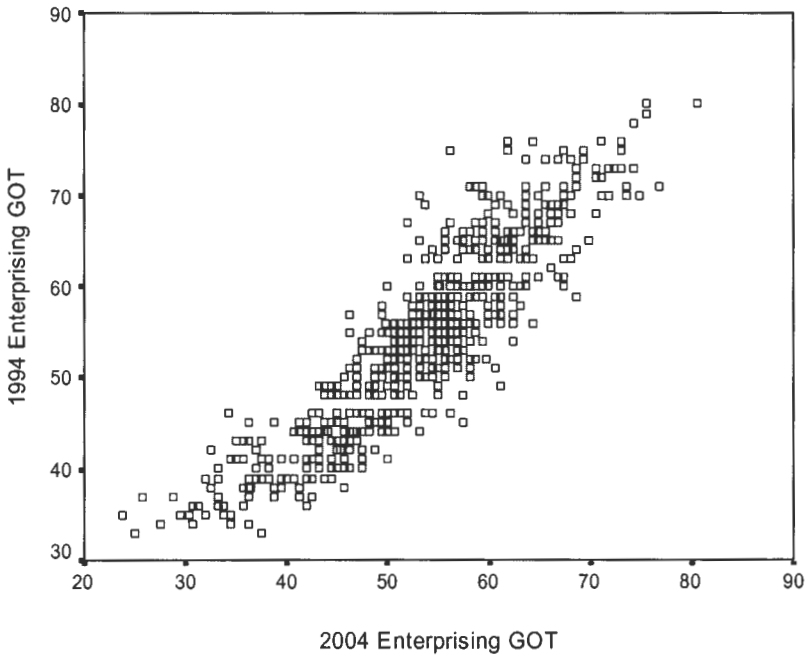
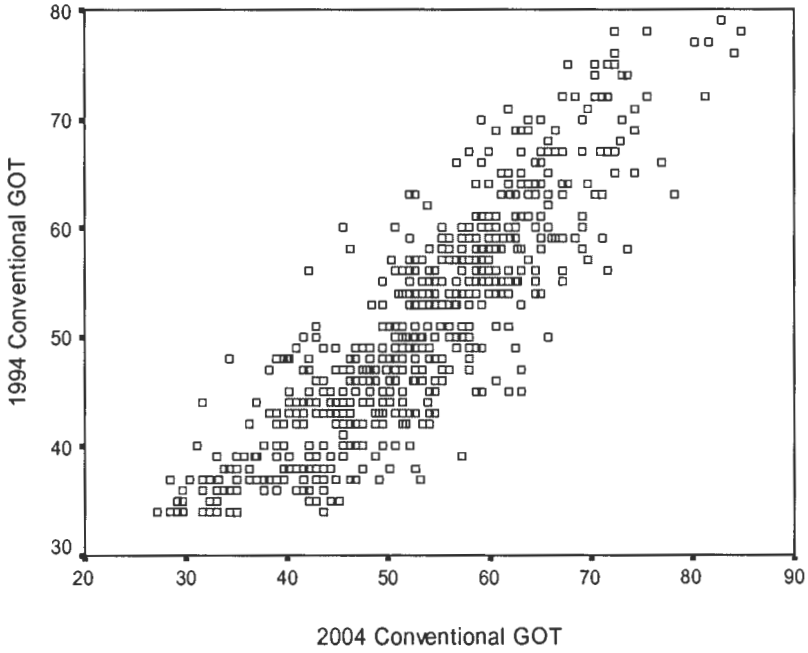


Figure 8. Scatterplot of 2004 and 1994 Conventional General Occupational Themes



Bivariate Correlations between Analogous Basic Interest Scales of the 1994 and 2004 Strong Interest Inventories by Sex

These Pearson product moment correlations tested the equivalency of participant scores on the analogous Basic Interest Scales of the 1994 and 2004 Strongs. Table 3.15 presents the correlations between each of the 22 analogous BISs of the 1994 and 2004 Strongs. All correlations were statistically significant at the $p < .001$ level.

Of the 44 total correlations, 10 failed to support the second hypothesis that all analogous content scales of the 1994 and 2004 Strongs would yield correlations above .85. These 10 relations (females listed first followed by males unless otherwise noted) included the following: 1) 1994 Agriculture and 2004 Nature and Agriculture (*rs* = .68; .64), 2) 1994 and 2004 Culinary Arts scales (*r* = .83; females only), 3) 1994 Social Service and 2004 Counseling and Helping (*rs* = .78; .78), 4) 1994 Organizational Management and 2004 Management (*r* = .84; males only), 5) 1994 Law/Politics and 2004 Politics and Public Speaking (*r* = .82; males only), 6) 1999 Law/Politics and 2004 Law (*rs* = .79; .80), and 7) 1999 Office Services and 2004 Office Management (*r* = .83; males only).

Table 3.17

Correlations between Analogous Basic Interest Scales of the 1994 and 2004 Strong Interest Inventories by Sex (n = 622)

1994 SII BISs	2004 SII BISs	Females (n = 374)	Males (n = 248)
Mechanical Activities	Mechanics & Construction	88	93

Table 3.17 (Continued)		**Females**	**Males**
Military Activities	Military	85	92
Nature	Nature & Agriculture	86	85
Agriculture	Nature & Agriculture	*68*	*64*
Athletics	Athletics	93	93
Science	Science	90	90
Medical Science	Medical Science	87	88
Mathematics	Mathematics	92	93
Applied Arts	Visual Arts & Design	88	90
Music/Dramatics	Performing Arts	91	91
Writing	Writing & Mass Communication	90	88
Culinary Arts	Culinary Arts	*83*	88
Social Service	Counseling & Helping	*78*	*78*
Teaching	Teaching & Education	91	88
Religion Activities	Religion & Spirituality	93	92
Medical Service	Healthcare Services	93	90
Sales	Sales	86	87
Organizational Mgmt	Management	85	*84*
Law/Politics	Politics & Public Speaking	86	*82*
Public Speaking	Politics & Public Speaking	89	97
Law/Politics	Law	*79*	*80*
Office Services	Office Management	87	*83*

Note. Decimals omitted. All parallel correlations are significant at the $p < .001$ level.

Correlations in italics denote r < .85

Bivariate Correlations between Analogous Personal Style Scales of the 1994 and 2004 Strong Interest Inventories by Sex

Table 3.18 presents the correlations between the four parallel PSSs of the 1994 and 2004 Strongs. All analogous PSSs were statistically significant at the $p < .01$ level. Of note is the Risk Taking scale, which displayed a correlation considerably lower than the expected .85 stated in the second hypothesis (*rs* = .74 and .71; females and males, respectively). Additionally, the male correlation between the 1994 and 2004 Learning Environment scales (listed above) also fell below the .85 threshold. Thus, three of the eight (38%) parallel PSS correlations failed to support the second hypothesis.

Table 3.18

Correlations between Analogous Personal Style Scales of the 1994 and 2004 Strong Interest Inventories by Sex (n = 622)

	Females (n = 374)				Males (n = 248)			
	2004 PSSs							
1994 PSSs	Work Style	Learning Environ.	Leader-ship	Risk Taking	Work Style	Learning Environ.	Leader-ship	Risk Taking
Work Style	**92**	05	44	07	**94**	09	52	04
Learning Environment	01	**88**	45	19	15	**79**	42	08
Leadership	47	51	**95**	26	56	47	**94**	20
Risk Taking	10	29	41	**74**	28	14	45	**71**

Note. Decimals omitted. All bold parallel correlations are significant at the $p < .01$ level.

CHAPTER 4. DISCUSSION

This section begins with a review of the preliminary analyses followed by a discussion of the two hypotheses. Within this review, major findings will be explored. This will be followed by a discussion of the limitations of this study including suggestions of additional areas in need of further research. Finally, the implications of this study and final conclusions are presented.

Preliminary Analyses

All three preliminary analyses conducted in this study will be discussed in turn. This section begins with a discussion of the normative comparisons between the current sample and both the 2004 GRS and Gasser samples. Each set of content scales is discussed individually. Next, attention is turned to the normalcy comparison of the 2004 and 1994 Strongs. Once again, the GOTs, BISs and PSSs are discussed in turn. Finally, the analysis of sex differences concludes the section.

Normative Comparisons of the Current Sample with 2004 GRS and Gasser Sample

Independent samples t-tests were run comparing the current sample of college students to the established norms of the 2004 Strong General Reference Sample of employed adults and to the Gasser sample (i.e., a subset of college samples taken from the 2004 GRS). A total of 82 comparisons were run for both females and males and a Bonferroni adjustment was made accordingly (.05 / 164 comparisons = .0003; $t = 3.62$).

General Occupational Themes. Table 3.1 above provides a good overview of the results of the normative comparisons made between all three samples for the GOTs. Several mean differences were found between samples, with only the Investigative GOT not differing significantly for either sex between any samples. For the Realistic, Social, and Enterprising

scales, all mean differences were statistically different between samples for both sexes ($p <$ .0003). Sex differences were observed for two of the GOTs with Artistic female means in the current sample differing significantly from the Gasser sample only, and the Conventional male means in the current sample differing significantly only from the GRS. Eta squared values were calculated for each scale in an effort to determine the effect size (and thus the meaning) of the observed differences between samples. Cohen's (1988) guidelines were used to reveal small effect sizes (eta squared = .01 to .03) for all but three of the observed differences between GOT means. Though mean differences were found between the current sample and the GRS and Gasser samples for five of the six GOTs, the pattern of small effect sizes speaks to the overall robustness of the GOTs to differentiate among the six Holland types. Additionally, these findings provide partial supportive evidence for the interpretative continuity of the GOTs from the 1994 to the 2004 Strong GOTs.

Basic Interest Scales. With their greater number and specificity, the BISs provide a richer source of information than the GOTs. Thus, an examination of the results of the mean differences between the BISs of the current sample, the 2004 GRS, and the Gasser sample are more revealing. Overall, females displayed fewer statistically significant mean differences ($p <$ .0003) than males (*ns* = 22 and 28, respectively). Additionally, a lesser number of statistically significant mean differences ($p <$ .0003) were observed between the current sample and the 2004 GRS than between the current sample and the Gasser sample (*ns* = 19 and 31, respectively). This is somewhat surprising considering both the current sample and the Gasser sample consist of college students while the 2004 GRS consists of employed adults. Finally, males also displayed an overall greater magnitude of difference between the means with 14 of the mean differences between male BISs falling above one-half a standard

deviation and only six of the mean differences between female BISs falling above one-half a standard deviation. This finding suggests that for males, the current sample was less representative of the 2004 GRS and the Gasser sample than for females, though the over-sampling of females for all three samples could partially account for the observed differences between sexes.

Personal Style Scales. The normative comparisons for the PSSs between the current sample and the 2004 GRS and Gasser sample revealed statistically significant mean differences ($p < .0003$) for only two PSSs, the Work Style scale and the Risk Taking scale. Sex differences were only observed for the Work Style scale with statistically significant female mean differences found between the current sample and both comparison samples, while statistically significant male differences were found only between the current sample and the GRS. These findings suggest that the PSSs in the current sample were closer representatives of the 2004 GRS and Gasser sample than either the GOTs or BISs. The only somewhat surprising finding is the lack of sex differences found on the Risk Taking scale as males have traditionally been found to score closer to the "Takes chances" pole of this scale than females (Donnay et al., 2005).

Normalcy Comparisons between the 1994 and 2004 Strong Interest Inventories

Shapiro-Wilks tests of normalcy were conducted to test for the normal distribution of each of the 35 content scales of the 1994 Strong and for each of the 41 content scales of the 2004 Strong by sex. The results of each set of content scales are described in turn.

General Occupational Themes. There was a noticeable difference between the 1994 and 2004 Strongs, as well as a difference in sexes for the 1994 Strong. All of the 2004 GOT tests were non-significant ($p > .002$) for both sexes indicating normal distribution for all

scales. However, seven of the 1994 GOTs were statistically significant ($p < .002$). Six of the seven statistically significant scales were the six 1994 female GOTs (with the exception of the male Artistic scale). The remaining five male GOTs were not statistically significant and thus denoted normal distribution. These findings would seem to indicate that the scores on the 2004 Strong were normally distributed across all the content scales as expected. However, all six female GOTs for the 1994 Strong being statistically significant with only one male GOT (Artistic) being statistically significant is interesting and worth noting, especially given the degree of item overlap between the two versions. One possible reason for the difference in results between the 1994 and 2004 Strongs could lie in the new five-point scoring index found on the 2004 Strong. The necessity of collapsing scores on the 1994 Strong in the current study in order to use its existing L-I-D scoring routine could account for some of these observed differences.

Basic Interest Scales. As stated previously, the similarity of the results of the GOTs and BISs are not surprising giving the hierarchical structure of the Strong. Once again, a noticeable difference occurred between the 1994 and 2004 Strongs, as well as a difference between females and males for both test versions. The majority of tests (41 / 50 = 82%) for the 1994 Strong resulted in statistical significance ($p < .0005$) indicating a violation of the normality assumption. Not surprising given the overall sex differences noted above, eight of the nine 1994 scales that were not statistically significant ($p > .0005$) occurred only for males. Again, following the pattern for the GOTs, the 2004 BISs showed almost opposite results from the 1994 BISs. Out of 60 tests, 45 (75%) failed to find statistical significance supporting normal distribution. Sex differences for the 2004 BISs were also consistent with 14 of the 15 statistically significant ($p < .0005$) BISs occurring only for females. These

results buttress the need for further investigation into sex differences, as well as expound the necessity of assessing differences due to the new five-point scoring index found on the 2004 Strong. Additionally, these observed sex differences could indicate a difference in female responses from the 1994 General Representative Sample to the 2004 GRS.

Personal Style Scales. The results of the normalcy tests on the PSSs also indicate differences between the 1994 and 2004 Strongs. Specifically, as with the other content scales, a greater percentage of the 2004 scales (90%) failed to find a Bonferroni adjusted statistical significance (.05 / 18 = .003) indicating normal distribution. Only the 2004 Strong male Risk Taking scale showed statistical significance ($p < .003$). For the 1994 Strong, half of the scales displayed statistical significance (50%) with three of the four being significant for females, and the fourth being significant for both males and females (i.e., 1994 Risk Taking scale). The further replication of the results found for the GOTs and BISs provide consistent support for the existence of sex differences on the 1994 Strong for this sample of respondents, as well as suggest an overarching difference in distributions between the 1994 and 2004 Strongs. Whether this difference is restricted to the current study due to its sampling methods (and resultant need to collapse the 1994 scores) or is more generalizable needs further investigation.

Univariate Analysis of Sex Differences

This analysis addressed the fourth of Smith and McCarthy's (1995) criteria by testing the ability of the 1994 and 2004 Strongs to discriminate between sexes. Additionally, it served as a test of the equivalency in magnitude of observed sex effects between the two versions. Univariate analyses of variance (ANOVAs) were conducted on the content scales of the 1994 and 2004 Strong Interest Inventories. Using Wilks's criterion, the examination of

the univariate effects demonstrated equivalent statistically significant differences between sexes. Results of these univariate analyses of variance (ANOVAs) for each of the content scales are discussed in turn.

General Occupational Themes. Each GOT univariate ANOVA was tested at the Bonferroni adjusted .004 level (.05 / 12 = .004). Unlike the tests of normality discussed earlier, the ANOVAs of between-subjects effects for sex differences were similar for both versions of the Strong. For both the 1994 and 2004 Strongs, statistically significant mean differences for sex were observed for the Realistic, Investigative, Artistic, and Social GOTs ($p < .001$). The Enterprising GOT did not show significant mean differences for sex on either the 1994 or the 2004 Strong ($p > .004$). Finally, the Conventional GOT resulted in a statistically significant mean sex difference only for the 2004 Strong.

Partial eta-squared values were used to examine the effect sizes of these observed differences with a strong effect found for both the 1994 and 2004 Realistic scales. Medium effect sizes were found for the 1994 Investigative, Artistic and Social GOTs, as well as the 2004 Social GOT. Small effect sizes were found for the 2004 Investigative, Artistic, and Conventional GOTs. These findings once again seem to indicate larger sex differences for the 1994 Strong versus the 2004 Strong as suggested by the analyses described above, though the magnitude of the differences between the two versions appears to be much less. That is, the pattern of both statistically significant and non-significant mean sex differences between the versions is fairly parallel indicating a similarity in sex differentiation between the two versions.

Basic Interest Scales. As with the GOTs, the results for the 1994 and 2004 Strongs were similar, as expected. In total, 17 of the 25 ANOVAs for the 1994 BISs were statistically

significant ($p < .001$), whereas 22 out of 30 of the 2004 BISs were statistically significant ($p < .001$). Non-significant results also showed a similar pattern between versions with six out of nine non-significant BISs being analogous between the two versions. Thus, the pattern of sex effects was similar for the BISs across both versions of the Strong, though three sets of analogous BISs displayed opposing results (i.e., 1994 Applied Arts and 2004 Visual Arts and Design, 1994 Teaching and 2004 Teaching and Education, and 1994 Sales and 2004 Sales).

Turning once more to Cohen's (1988) criteria, an examination of the partial eta-squared values revealed large effect sizes for three of the 1994 Strong and four of the 2004 BISs, medium effect sizes for nine of the 1994 and for eight of the 2004 BISs, and small effect sizes for four of the 1994 BISs and for nine of the 2004 BISs. As stated previously, due to the structural design of the Strong, the more specific BISs often reveal more in-depth information than the GOTs. In the case of this analysis, the BISs cement the similarity of mean sex differences observed between the 1994 and 2004 Strongs. Further, the results suggest similar differential interpretation based on sex for both versions, with caution taken for the opposing results of the three analogous BISs pending further investigation.

Personal Style Scales. The tests of between-subjects sex effects for the PSSs actually showed the most similarity between the 1994 and 2004 Strongs. Specifically, the univariate ANOVAs were statistically significant ($p < .006$) for the Work Style and Risk Taking PSSs, and statistically non-significant ($p > .006$) for the Learning Environment and Leadership PSSs for both the 1994 and 2004 Strongs. Additionally, effect sizes were parallel for both versions with the Work Style and Risk Taking scales having large effect sizes (Cohen, 1988) and all remaining scales for both versions showing small effect sizes. The results of the PSSs

provide the greatest support for the similarity of mean sex differences between the 1994 and 2004 Strongs.

The import of the close parallel between the old and new versions of the Strong lies in the ability of the user to trust in the results of the updated 2004 version. The resulting similarity of the 1994 and 2004 content scales to demonstrate sex effects with comparable effect sizes provides evidence of the equivalence of sex differences between the two versions. This helps instill needed confidence in the similarity of interpretation for both sexes using either the 1994 or the 2004 Strong, and thus serves to increase user acceptance as recommended by Campbell (1972). This analysis attempted to partially address Smith and McCarthy's (1995) fourth criterion for test calibration, demonstrating item discriminatory power. The results of the univariate tests allowed for the demonstration of the ability of both versions of the Strong to discriminate fairly equally between sex. However, much more research needs to be conducted to allow for the full exploration of this criterion. This is discussed further under future directions.

Hypotheses

This section discusses the implications of the results of the two main hypotheses tested in this study. First, the hypothesis testing the equivalence of the parallel content scales of the 1994 and 2004 Strongs will be discussed. This is followed by a discussion of the implications of the results of the bivariate correlations between the analogous content scales of the two versions of the Strong.

Hypothesis One

The first main hypothesis presented an attempt to demonstrate the equivalency of the parallel content scales of the 1994 and 2004 Strongs. To accomplish this, paired sample t-

tests were used to explore any statistically significant differences between the means of the parallel content scales of the 1994 and 2004 Strongs. Specifically, no mean differences greater than one-half standard deviation were expected, with one standard deviation equaling 10 points. Each set of content scales will be discussed in turn.

General Occupational Themes. All the t-tests on the GOT means supported this first hypothesis that no statistically significant mean differences greater than one-half a standard deviation would exist between parallel content scales of the 1994 and 2004 Strongs. Specifically, the largest mean difference found was between the Realistic means for females (mean difference = 5.66). These findings indicate that for the analogous GOTs, mean scores between the 1994 and 2004 would not differ by more than one-half a standard deviation.

Basic Interest Scales. Seven of the BIS t-tests failed to find support for this first hypothesis. As seen in table 3.13, of these seven, six were statistically significant ($p < .001$) for females but not males. Interestingly, the six statistically mean differences for females fell across the RIASEC (i.e., two Realistic BISs, one Investigative BIS, one Artistic BIS, and one Social BIS). Additionally, all statistically significant mean differences resulted in higher mean scores on the 2004 Strong than on the 1994 Strong. These differences are even more impressive given the large number of identical items on the two versions in this study and suggest several implications.

It seems that women are scoring higher (and closer to the men) on the 2004 version of the Strong. This could be due to the existence of differences in the norming samples of the 1994 and 2004 Strongs. For example, perhaps additional Realistic items were geared toward women in the restandardization of the 2004 Strong, partially explaining their higher mean scores observed above. Differences in item wording and the deletion of outdated items could

also account for the marginal rise in 2004 Social, Enterprising, and Conventional GOT scores for women. This could indicate more reliable measurement of female interests in the 2004 Strong. These results suggest the need for the conduction of future research using factor analytic methodology. Implications for profile interpretation due to these observed mean differences are provided under Implications for Future Counselors.

Personal Style Scales. Of the eight paired sample t-tests run on the PSSs by sex, five resulted in a statistically significant difference between mean scores on the 1994 versus the 2004 Strong. Unlike with the BISs, these differences were equally spread across sex with three female PSSs and two male PSSs showing significance. As with the GOTs, all significant mean differences fell below one-half a standard deviation as thus supported the first hypothesis. These results are not unexpected as the four analogous PSSs underwent fewer changes than the BISs.

This first hypothesis attempted to address Smith and McCarthy's (1995) third criterion, the content homogeneity of underlying facets to ensure equal construct representation. Though paired-samples t-tests can certainly not speak to the construct validity of the 2004 Strong, they do provide a method of determining if underlying facets (i.e., content scales) of one version of a test are measuring the same thing as a second version (albeit a simplistic method). Again, it is highly recommended that factor analytic methods be used to further investigate the underlying facets and hierarchical structure of the 2004 Strong.

Hypothesis Two

For the second hypothesis, the author calibrated the 1994 and 2004 Strongs by using Pearson product moment correlations to demonstrate the equivalency of the hierarchical conceptual structure (Smith and McCarthy's [1995] first criterion) of the two versions.

Additionally, this second hypothesis attempted to address the additional purpose of this study, which is to determine the interpretive implications of the 2004 revision.

General Occupational Themes. The results of the GOT correlations between the 1994 and 2004 Strongs provide support for the second hypothesis and display no evidence of meaningful differences between parallel scales. Specifically, all parallel GOT correlations were above .85 as expected. The reader is referred to Figures 3 through 7 which present the scatterplots of the parallel GOT and display the strong positive correlations between the 1994 and 2004 Strong GOTs.

Basic Interest Scales. Of the 44 total correlations conducted on the BISs, only 10 failed to support this second hypothesis by failing to yield correlations above .85. The highest correlations were found between the Athletics scales (for both sexes), the Mathematics scales (for both sexes), the 1994 Religion Activities and 2004 Religion and Spirituality scales (for both sexes), and the 1994 Public Speaking and 2004 Politics and Public Speaking scales (for males). The lowest correlations were found between the 1994 Agriculture and 2004 Nature and Agriculture scale (for both sexes), the Culinary Arts scales (for females only), the 1994 Social Service and 2004 Counseling and Helping scales (for both sexes), and the 1994 Law/Politics and 2004 Law scale (for both sexes).

These findings are somewhat surprising given the large number of items that were responded to only once, but that were scored on both versions of the Strong in this sample. Of special interest is the correlations between the 1994 Agriculture and 2004 Nature and Agriculture scale (*rs* = 68 and 64 for females and males, respectively) and between the 1994 Nature and 2004 Nature and Agriculture scale (*rs* = 86 and 85 for females and males, respectively). Given the relatively low strength of these correlations, the decision to combine

these scales in the 2004 Strong becomes questionable. Further evidence is provided in Staggs (2002), who conducted a meta-analyses of the convergence of interests and personality using the 1994 Strong and the Multidimensional Personality Questionnaire (Tellegen, 1982). In this study, personality differences were found with the 1994 Strong Nature scale positively relating to Absorption and with the 1994 Strong Agriculture scale negatively relating to Harm Avoidance suggesting qualitative differences between the 1994 Nature and Agriculture scales.

Personal Style Scales. All parallel PSSs were statistically significant at the $p < .01$ level. However, three out of eight (38%) of the PSS scales failed to support this second hypothesis. Both the Risk Taking PSS (for both sexes) and the Learning Environment PSS (males only) resulted in correlations lower than the expected .85. Again, though the majority of the correlations all fell above .85 as expected, 38% failing to meet this criteria is a possible cause of concern for test administrators, developers, and counselors alike. Unfortunately, it was not possible to take into account the effect of the new 2004 Team Orientation scale in this study which could partially explain the observed lowered correlations.

Strong positive relations were found between the overwhelming majority of the analogous content scales of the 1994 and 2004 Strongs. This finding is arguably the most important in this study as determining how well the content scales across the 1994 and 2004 Strongs calibrate ensures that the interpretation of the scales for the 2004 version closely approximate the interpretation of the scales using the 1994 version. Millions of Strong users are very knowledgeable about the interpretive guidelines of the 1994 Strong, if the new 2004 version is to be accepted; studies must demonstrate similar interpretive guidelines for the 2004 Strong (Campbell, 1972). The strong positive correlations between most of the

analogous content scales (especially between the GOTs) successfully addresses both user concerns regarding the stability of the Strong's theoretical structure and, consequently, Smith and McCarthy's (1995) first criterion. However, the small minority of correlations falling below the .85 threshold should not be ignored.

Limitations

This study had several limitations. Most of the limitations stem from sampling procedures used to procure the data. To begin, because the research version of the 2004 Strong was used, participants were forced to respond to all items on the new five-point item response format. This necessitated the need to collapse the 1994 Strong items into their traditional three-point LID response format to facilitate scoring. This limitation was specifically addressed by a similar study as the current study, in which the continuity of the 1994 and 2004 Strongs was examined (Borgen et al., 2004). In this study, participants were given both the 1994 and 2004 Strong testing booklets, thus avoiding item replication.

Second, because the data was collected from an upper Midwest university, the majority of the sample was Caucasian. As with the limitation mentioned above, this serves to decrease the overall generalizability of the findings in this study. Finally, due in part to the propriety nature of the Strong's scoring routines and in part to the nature of this study, this study was unable to conduct more sophisticated statistical analyses capable of more directly addressing Smith and McCarthy's (1995) five criteria of test calibration such as discriminant function analysis. Some suggestions for future analyses are presented in the next section.

Future Directions for Research

This study presents one of the initial attempts to establish the reliability of the 2004 Strong by calibrating it with its 1994 predecessor. The 2004 Strong is in its infancy, and as

such much more work needs to be done demonstrating its reliability and validity. It is particularly necessary to be prudent about addressing the many theoretical concerns and empirical lessons inherent in the process of test calibration (Adams, 2000; Campbell, 1972; Reise, Waller, & Comrey, 2000; Smith and McCarthy, 1995). Future researchers and test developers should heed these warnings and attempt to address these concerns along each step of the test revision process. For example, internal consistency reliability needs to be established directly addressing Smith and McCarthy's second criterion. Additional item level analyses must also be conducted in order to determine the stability of the hierarchical structure of the Strong from the 1994 to the 2004 version which would more fully address Smith and McCarthy's first criterion (i.e., the calibration of the hierarchical structure). Finally, factor analyses must be conducted in order to demonstrate the factor replication of the 1994 Strong in the 2004 Strong (addressing Smith and McCarthy's fifth criterion).

Implications for Counselors

The second purpose of this study was to move beyond mere test calibration in order to help determine the interpretive implications of the 2004 Strong revision. The prompting of this additional purpose was twofold. First, it was hoped to aid in the establishment of user acceptance as recommended by Campbell (1972). Second, the primary purpose of the Strong is to serve as a tool in career counseling. Millions of Strongs are given each year and the results often help shape and/or reshape the career paths of thousands upon thousands of people each year. It is therefore imperative that counselors be able to trust in the new 2004 version of the Strong. That is, counselors need to be able to transfer the knowledge (and thereby comfort) they hold about the 1994 Strong to the new 2004 version.

The results of this study have generally supported the successful calibration of the two versions, with several implications for counselors having been raised pending further research. First, the ramifications of the new 2004 five-point response format on the distribution of scores are unclear. It is possible that scores could fall further into the extremes causing very high or very low scores to become more common on the 2004 Strong.

Second, the observed mean sex differences between analogous scales of the 1994 and 2004 could have specific implications for women. Because the observed mean differences on the content scales of the two versions resulted in females scoring higher on some of the 2004 BIS scales, their interest levels on some of the GOTs (particularly the Realistic GOT) could become inflated compared to how they would score on the 1994 Strong. Counselors will have to be cautious in how they interpret these higher scores as they may be more reflective of differences in the standardization samples rather than reflections of true individual differences.

Finally, though the majority of the correlations between the analogous scales of the 1994 and 2004 Strongs showed a strong positive relation, not all scales were as strongly correlated as expected. Thus, once again counselors should use caution in the direct interpretation of the 2004 Strong using 1994 interpretative strategies, and ensure that more extreme scores are not over-interpreted (especially for females).

Conclusion

This study provided a good initial attempt at test calibration between the 1994 and 2004 Strong Interest Inventories, and demonstrated the overall stability of the interpretive implications of the 2004 Strong. One of the main purposes of this study was to heed

Campbell's (1972) warning that any new revision must gain user acceptance or risk having users continue to rely on the older, more familiar version.

Numerous questions remain to be answered before the 1994 and 2004 Strongs it can be stated with any confidence that the two versions are completely calibrated. Determining the interpretative equivalency of two versions of any psychometric instrument is an arduous task, and the Strong is no exception. Additional research must be conducted on the 2004 Strong to further inform its users of the measure's reliability and validity. The popularity of the Strong rests in part on its theoretical strength and in part on its empirical strength. Both of these areas require further investigation before the 2004 Strong can be said to be an improved, but fundamentally unchanged, version of the 1994 Strong. In conclusion, though more work will be needed before many users' comfort threshold is reached; it hoped that the results of this study serve as a beginning for the incremental increase in user acceptance of the new 2004 Strong Interest Inventory.

CHAPTER 5. REFERENCES

Adams, K. M. (2000). Practical and ethical issues pertaining to test revisions. *Psychological Assessment, 12,* 281-286.

Allport, G. W., Vernon, P. E., & Lindzey, G. (1970). *Study of Values* (3rd Ed). Iowa City, IA: Riverside Publishing Company.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Betz, N. E., Borgen, F. H., & Harmon, L. W. (1996). *Skills Confidence Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Betz, N., Borgen, F., Rottinghaus, P., Paulsen, A., Halper, C. & Harmon, L. (2003). The Expanded Skills Confidence Inventory: Measuring basic dimensions of vocational activity. *Journal of Vocational Behavior, 62,* 76-100.

Betz, N. E., & Taylor, K. M. (1982). Concurrent validity of the Strong-Campbell Interest Inventory for graduate students in counseling. *Journal of Counseling Psychology, 29* (6), 626-635.

Borgen, F. H. (1972). Predicting career choices of able college men from Occupational and Basic Interest Scales of the Strong Vocational Interest Blank. *Journal of Counseling Psychology, 19,* 202-211.

Borgen, F. H. (1986). New approaches to the assessment of interests. In B. W. Walsh & S. H. Osipow (Eds.), *Advances in vocational psychology, Vol. 1: The assessment of interests* (pp. 83-125). Hillsdale, NJ: Erlbaum.

Borgen, F. H. (in press). Vocational interests. In C Speilberger (Ed.), *Encylopedia of Applied Psychology*. San Diego, CA: Academic Press.

Borgen, F. H., Larson, L. M., Bailey, D. C., Thompson, R. C., & Donnay, D. A. C. (2004). *Continuity of the 1994 and 2004 Strong Interest Inventories.* Manuscript submitted for publication.

Cairo, P. C. (1979). The validity of the Holland and Basic Interest scales of the Strong Vocational Interest Blank: Leisure activities versus Occupational membership as criteria. *Journal of Vocational Behavior, 15,* 68-77.

Campbell, D. P. (1966). *Strong Vocational Interest Blanks manual*. Stanford, CA: Stanford University Press.

Campbell, D. P. (1969). *Strong Vocational Interest Blanks manual supplement.* Stanford, CA: Stanford University Press.

Campbell, D. P. (1971). *Handbook for the Strong Vocational Interest Blank.* Palo, Alto, CA: Stanford University Press.

Campbell, D. P. (1972). The practical problems of revising an established psychological test. In J. N. Butcher (Ed.), *Objective personality assessment: Changing perspectives* (pp. 117-130). New York: Academic Press.

Campbell, D. P. (1974). *Manual for the Strong-Campbell Interest Inventory.* Stanford, CA: Stanford University Press.

Campbell, D. P., & Borgen, F. H. (1999). Holland's theory and the development of interest inventories. *Journal of Vocational Behavior, 55,* 86-101.

Campbell, D. P., Borgen, F. H., Eastes, S., Johansson, C. B., & Peterson, R. A. (1968). A Set of Basic Interest Scales for the Strong Vocational Interest Blank for men. *Journal of Applied Psychology Monographs, 52* (6, Part 2).

Campbell, D. P., & Holland, J. L. (1972). A merger in vocational interest research: Applying Holland's theory to Strong's data. *Journal of Vocational behavior, 2,* 353-376.

Campbell, D. P., & Hansen, J. C. (1981). *Manual for the Strong-Campbell Interest Inventory* (3rd ed.). Stanford, CA: Stanford University Press.

Campbell, D. P., Hyne, S. A., & Nilsen, D. L. (1992). *Manual, Campbell Interest and Skills Survey*. Minneapolis, Minnesota: National Computer Systems.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dolliver, R. H. (1975). Concurrent prediction from the Strong Vocational Interest Blank. *Journal of Counseling Psychology, 22,* 199-203.

Donnay, D. A. C., & Borgen, F. H. (1996). Validity, structure, and content of the 1994 Strong Interest Inventory. *Journal of Counseling Psychology, 43,* 275-291.

Donnay, D. A. C., Morris, M., Schaubhut, N. & Thompson, R. (2005). *Strong Interest Inventory manual: Research, development, and strategies for interpretation*. Palo Alto, CA: Consulting Psychologists Press.

Gasser, C. E. (2005). Concurrent validity of the 2004 Strong Interest Inventory: An examination of sex and college major. Unpublished doctoral dissertation, Iowa State University.

Hansen, J. C. (1986). Strong Vocational Interest Blank/Strong-Campbell Interest Inventory. In W. B. Walsh & S. H. Osipow (Eds.). *Advances in vocational psychology* (pp. 1-29). Hillsdale, NJ: Erlbaum.

Hansen, J. C. (1990). Edward Kellogg Strong, Jr.: First author of the Strong Interest Inventory. In P. P. Heppner, (Ed.), *Pioneers in counseling and development: Personal and professional perspectives.* Alexandria, VA: American Association for Counseling and Development.

Hansen, J. C., & Campbell, D. C. (1985). *Manual for the Strong Interest Inventory 4th ed. Stanford,* CA: Stanford University Press.

Hansen, J. C., Collins, R. C., Swanson, J. L., & Fouad, N. A. (1993). Sex differences in the structure of interests. *Journal of Vocational Behavior, 42,* 200-211.

Hansen, J. C., & Johansson, C. B. (1972). The application of Holland's vocational model to the Strong Vocational Interest Blank for Women. *Journal of Vocational Behavior, 2,* 479-493.

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

Hathaway, S. R., & McKinley, J. C. (1940). The MMPI manual. New York: Psychological Corporation.

Holland, J. L. (1966). *The psychology of vocational choice.* Waltham, MA: Blaisdell.

Holland, J. L. (1985). *Making vocational choices* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Holland, J. L., & Gottfredson, G. D. (1976). Sex differences, item revisions, validity, and the Self-Directed Search. *Measurement and Evaluation in Guidance, 8,* 224-228.

Holland, J.L. (1994). *The self-directed search: Professional users guide.* Odessa FL: Psychological Assessment Resources.

Isaacs, J., Borgen, F. H., Donnay, D. A. C., & Hansen, T. A. (1997). *Self-efficacy and interests: Relationships of Holland themes to college major.* Poster presentation at the 105th Annual convention of the American Psychological Association, Chicago.

Jackson, D. N. (1977). *Jackson Vocational Interest Survey manual.* Port Huron, MI: Research Psychologists Press, Inc.

Johansson, C. B. (1986). *Career Assessment Inventory: The enhanced version.* Minneapolis, NM: National Computer Systems.

Johnson, R. W., & Johnson, C. B. (1972). Moderating effect of basic interests on predictive validity of the SVIB occupational scales. *Proceedings, 80th Annual Convention. American Psychological Association,* pp. 589-590.

Kuder, F. & Zytowski, D.G. (1991). *Kuder occupational interest survey, General manual (3rd ed.).* Adel, IA: National Career Assessment Services, Inc.

Lattimore, R. R., & Borgen, F. H. (1999). Validity of the 1994 Strong Interest Inventory with racial and ethic groups in the United States. *Journal of Counseling Psychology, 46(2),* 185-195.

McArthur, C. (1954). Long-term validity of the Strong interest test in two subcultures. *Journal of Applied Psychology, 38,* 346-353.

Olsen, L. L. (1996). Concurrent validity of the 1994 Strong Interest Inventory: A comparison of criterion groups by sex. Unpublished doctoral dissertation, Iowa State University.

Prince, J. P., & Heiser, L. J. (2000). Essentials of Career Interest Assessment. New York, NY: John Wiley & Sons, Inc.

Ralston, C. A., Borgen, F. H., Rottinghaus, P J., & Donnay, D. A. C. (in press). Specificity in interest measurement: Basic Interest Scales and major field of study. *Journal of Vocational Behavior.*

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12,* 287-297.

Rottinghaus, P. J., Lindley, L. D., Green, M. A., & Borgen, F. H. (2002). Educational aspirations: The contribution of self-efficacy, interests, and personality. *Journal of Vocational Behavior, 61,* 1-19.

Savickas, M. L., Taber, B. J., & Spokane, A. R. (2002). Convergent and discriminant validity of five interest inventories. *Journal of Vocational Behavior, 61,* 139-184.

Schaubhut, N. A., Donnay, D. A. C., Gasser, C. E., & Borgen, F. H. (2004). *Validity of 2004 Strong Interest Inventory: Sex and ethnicity effects.* Poster proposal for the 112th Annual convention of the American Psychological Association, Honolulu.

Schmidt, D. B., Lubinski, D., & Benbow, C. P. (1998). Validity of assessing educational-vocational preference dimensions among intellectually talented 13-year olds. *Journal of Counseling Psychology, 45 (4),* 436-453.

Shapiro, S. S. and Wilks, M. B. (1965). An analysis of variance test for normality (complete samples), *Biometrika, 52, 3 and 4,* pages 591-611.

Smith, G. T., & McCarthy, D. M. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment, 7,* 300-308.

Spokane, A. R. (1979). Validity of the Holland categories for college women and men. *Journal of College Student Personnel, 20,* 335-340.

Staggs, G. D. (2002). Meta-analyses of interest-personality convergence using the Strong Interest Inventory and the Multidimensional Personality Questionnaire. Unpublished master's thesis, Iowa State University.

Strauss, E., Spreen, O., & Hunter, M. (2000). Implications of test revisions for research. *Psychological Assessment, 12,* 237-244.

Strong, E. K. Jr. (1927). *Vocational Interest Blank.* Stanford, CA: Stanford University Press.

Strong, E. K., Jr. (1943). *The vocational interests of men and women.* Stanford, CA: Stanford University Press.

Strong, E. K. Jr. (1955). *Vocational interests 18 years after college*. Minneapolis: University of Minnesota Press.

Swaney, K.B. (1995). *Technical manual: Revised Unisex Edition of the ACT Interest Inventory (UNIACT)*. Iowa City, IA: American College Testing.

Tellegen, A. (1982). *Brief Manual for the Differential Personality Questionnaire*. Unpublished manuscript, University of Minnesota.

Tilton, J. W. (1937). The measurement of over-lapping. *Journal of Educational Psychology, 28,* 656-662.

Tracey, T. J., & Hansen, J. C. (1993). Evaluating Holland's and Gati's vocational-interest models: A structural meta-analysis. *Psychological Bulletin, 113.* 229-246.

Tracey, T. J., & Rounds, J. B. (1993). Evaluating Holland's and Gati's vocational-interest models: A structural meta-analysis. *Psychological Bulletin, 113,* 229-246.

Tulsky, D., Zhu, J., & Ledbetter, M. F. (1997). *WAIS-III/WMS-III technical manual.* San Antonio, TX: The Psychological Corporation.

Varca, P. E., & Shaffer, G. S. (1982). Holland's theory: Stability of avocational interests. *Journal of Vocational Behavior, 21,* 288-298.

Wechsler, D. (1981). *WAIS-R manual.* New York: The Psychological Corporation

Worthington, E. L., & Dolliver, R. H. (1977). Validity studies of the Strong Vocational Interest Inventories. *Journal of Counseling Psychology, 24,* 208-216.

Zytowski, D. G. (1985). *Manual supplement for the Kuder Occupational Interest Survey.* Chicago, IL: Science Research Associates.

Zytowski, D. G., & Warman, R. E. (1982). The changing use of tests in counseling. *Measurement and Evaluation in Guidance, 15,* 147-152.