

Gender Differences in the Validity of Career Interest Inventories

Stephanie T. Burns

Counselor Education and Counseling Psychology, Western Michigan University, Kalamazoo, USA
Email: stephanie.burns@wmich.edu

Received 29 March 2014; revised 25 April 2014; accepted 21 May 2014

Copyright © 2014 by author and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Predictive validity (including hit rates, kappa coefficients, and chance expectancy rates) between standard scoring and person matching was compared by gender based upon ex post facto data collected on 5143 medical students who had taken a career interest inventory and entered their medical residency. Hit rate accuracy for person matching with females and males in this study was lower than standard scoring. However, person matching demonstrated greater gender balancing in first match hit rates. Additionally, person matching increased career interest inventory validity over standard scoring as it has the greater ability to a) differentiate between and b) assign to specific occupational groups for females and males. Furthermore, person matching has the potential to offer female and male test takers the ability to receive narrative career data, which could improve the career decision making process over the scoring reports of career interest inventories using standard scoring.

Keywords

Person Matching, Predictive, Hit Rate, Gender, Career Interest

1. Introduction

Career interest inventories are utilized annually by thousands of people in the US as the foundation for making career decisions (Donnay, 1997). Research has demonstrated that gender differences are minimal to absent for most psychological variables in the US (large gender difference exceptions include motor behaviors such as throwing distance and some aspects of sexuality; moderate gender differences in aggression) (Hyde, 2005). However, concerns exist that career interest inventories demonstrate gender differences in hit rates (Su et al., 2009; Hackett & Lonborg, 1993; Wetzel et al., 2012; Einarsdóttir & Rounds, 2009). Hit rates are operationally defined as an exact match between the individual's chosen occupation and the occupation suggested by the psy-

chometric scoring methodology. The greater the accuracy with which a psychometric scoring methodology calculates participants' final occupational choice, the higher the hit rate reported. Further, studies of interest inventories have revealed that each gender resembles others of the same gender in a different occupation more than members of the opposite gender in the same occupation (Kuder, 1977a; Whiston & Bouwkamp, 2003). As such, there has remained a considerable gender bias in career interest inventory development and scoring.

As the interest is influenced greatly by parents' expectations, societal values, culture, and the child's exposure to permissible activities, gender differences play an important role in career interest inventory development and occupational selection (Su et al., 2009; Harmon, 1997; Li & Kerpelman, 2007). Strong (1943) noted that gender differences in interest are apparent by 15 years of age and are never unlearned. In general, men prefer to work with things over people and women prefer to work with people over things (Su et al., 2009; Strong, 1943; Lippa, 1998; Stockdale & Nadler, 2012). Strong's (1943) research suggested that women's interests appear to be comprehensive and the heterogeneity in female interests makes it harder to specify occupational interest patterns for women. In addition, the duties of the average female worker are not the same as those of the average male worker in the same occupation (Kuder, 1977a; Strong, 1943; Stockdale & Nadler, 2012). Further creating a gap, career interest inventories may be initially designed for testing men with the later inclusion of women into the already developed instrument (Kuder, 1977a; Harmon, 1997; Strong, 1943).

A meta-analysis by Su, Rounds, and Armstrong (2009) suggests that gender differences in career interest inventories are still a concern as females align with an interest in people and prefer social, artistic, and conventional interest categories and males align with an interest in things and prefer realistic and investigative categories. These differences between the genders remain consistent despite changes in age and life-span development. Ultimately, gender differences explain trends in females refraining from occupations in science, technology, engineering, and mathematics in the US (Su et al., 2009; Stockdale & Nadler, 2012).

Current research has considered possibilities for reducing gender differences in career interest inventories, which can be detected at the item and scale levels (Su et al., 2009). Research exploring gender balancing in career interest inventories suggests incorporating only those items that have little to no disparity in response rates between the genders (Whiston & Bouwkamp, 2003). However, this approach has led to interest inventories with reduced construct validity rates and hence a still unsettled debate about society's needs versus construct validity in the development of career interest inventories (Su et al., 2009).

2. Career Interest Scoring Traditions

Three psychometric traditions have dominated the field of career interest inventory construction. Strong's empirically-based group criterion related measurement of interest, Holland's theoretically-based prototype criterion related measurement of interest, and Kuder's person matching-based measurement of interest serve as the foundation of interest measurement (Donnay, 1997). Strong theorized that creating occupational scales linked the test taker to real world behavior in that occupation (Campbell & Borgen, 1999). To create an occupational scale, interest inventory scores from 250 to 500 individuals in an occupation are compared with the scores of nearly 5000 men-in-general and women-in-general (Kuder, 1977a; Strong, 1943; Campbell & Borgen, 1999). Using discriminant function analysis a small number of items that show large differences would be assigned beta weights to determine how much the item closely resembled or differed significantly from the preferences of the occupational group. The small numbers of items ranked by the test taker are then adjusted up or down using beta weights for each of the occupations measured by the career interest inventory. The final score for the scales could classify an individual as a member of this or that occupational group by factoring out common interests (Kuder, 1977a; Strong, 1943; Campbell & Borgen, 1999). Initial research suggested that for interest inventories to offer meaningful results to both genders there must be women's scales for women and men's scales for men as combined gender scales were less valid predictors of occupational choice than creating one single-gendered scale for an occupation (Donnay, 1997; Kuder, 1977a; Strong, 1943). While the need for different scales for each gender remains today when using Strong's theory, not every career interest inventory creates separate norms for the genders, which can increase gender imbalances in the hit rates of career interest inventories (Donnay, 1997).

As a classification interviewer in the Army, Holland noted that people tended to exemplify one of only a few vocational personality types and came to the conclusion that vocational interest was an aspect of the personality (Nauta, 2010; Holland, 1966). In 1959, Holland proposed an occupational classification system comprised of six

categories: Realistic, Intellectual, Artistic, Social, Enterprising, and Conventional (Holland et al., 1969). From this, he developed a subtle coding system to describe a typical prototype for both persons and environments based upon initials from the six scales (Holland et al., 1969; Holland, 1961). A Holland code (which is created using the initials of the highest three scoring scales in order of most to least resemblance) allows for 720 different variations of personality (Holland, 1966). To empirically validate the six categories, a factor analysis was performed in 1968, which suggested that each of the six scales measured something different from the other six scales and that there are at a minimum six different kinds of people (Holland et al., 1969). All six of his scales were statistically significant for men, yet women's data yielded only four scales: Intellectual, Artistic, Social, and Conventional (Holland et al., 1969). Holland's ability to assess both individuals and work environments via stereotypical examples (or prototypes) offers a parallel way to link the two together and increases the effectiveness of career interventions (Nauta, 2010; Holland, 1966).

Kuder wanted clients to take a career interest inventory and then be matched to the narrative autobiography of individuals who had answered the interest inventory in a similar manner. His theory of career interest stated that women especially would benefit from the increased knowledge that would be gained from receiving autobiographical occupational information in addition to naming specific occupations that the individual resembled (Kuder, 1977a). Kuder called this technique person matching and it eliminated the problem of creating occupational scales based upon a small group of individuals from an occupation, which was the basis of Strong's work and it ended trying to make people fit a stereotypic framework as was the case with Holland's work (Kuder, 1977a). Kuder created a reference group of at least 5000 people from several hundred occupations, which was not possible through the use of Strong's occupational group scales. Even today inventories using Strong's theory must limit the number of occupations included in the scoring report. For example, the Campbell Interest and Skill Survey, which is scored using Strong's theory, can only calculate scores for 60 different occupations when there are over 800 occupations in the US today. Further, Kuder scored all items on the interest inventory to measure similarities as well as differences (Donnay, 1997; Ihle-Helledy et al., 2004). Using person matching, all of the test taker's scores on all items were compared to all members of the reference group to offer the test taker autobiographies (including current occupations, past occupations, lifestyles, future goals, and descriptions of what they like best and least about their occupation) of the closest 20 reference group members (Kuder, 1977a). The person matching scoring report allowed the test taker to reflect on the career themes found within the narratives to improve career decision making processes (Kuder, 1977b).

Kuder suggested several benefits to person matching. Person matching could overcome the problem of gender bias found in occupational groups or prototypes since the person matching reference group includes members of both genders equally in a wider variety of occupations (Kuder, 1977b). Additionally, new occupations could become part of the career interest inventory as soon as the data from a few people in the occupation are collected without waiting for a large number of cases from which to build a new occupational scale or measure the stereotypic personality type of the individuals who comprise the new occupation (Kuder, 1977b). Moreover, person matching does not assume stable occupations, which is important in a global economy demanding flexibility and evolution in the career paths of individuals due to more outsourcing and contractual work.

In addition to gender difference issues, researchers are challenged with incorporating current vocational constructions into the theory and science behind career interest inventories (Armstrong & Rounds, 2010). Postmodern and constructivist philosophies are increasingly becoming the foundation of career counseling (Amundson, 2009; Cochran, 1997; McMahon & Watson, 2010; Peavy, 1998; Savickas et al., 2009). This shift emphasizes the use of narratives and story to activate meaning making processes to assist the client in constructing their desired future. In general, current postmodern and constructivist career counseling interventions are administered by trained counselors, which limit their reach. Career interest inventories currently lack significant career meaning making processes for clients who are making career decisions through the use of narratives and story.

For this study, gender differences in career interest inventory validity were investigated between the empirical method of Strong (developed in 1927) and the person matching method of Kuder (developed in 1977). Career interest inventories have two measures of validity (Strong, 1943). First, validity rests on an interest inventory's ability to differentiate between specific occupational groups (Kuder, 1977a; Strong, 1943). Second, the interest inventory needs to assign individuals to membership in one or more occupational groups based upon their interest inventory scores (Strong, 1943). A part of measuring the assignment of individuals to an occupational group includes calculating hit rates.

To make a comparison between the two methods, the career interest inventory selected needed to include a

large sample with item scores, longitudinal data, and demographic data. While any profession could have been utilized to empirically examine gender differences, medicine was used because of the many specialty choice problems that arise for medical students. Medicine has over 100 medical specialties to choose from, which requires decidedly different abilities, skills, and talents (Sodano & Richard, 2009; Rogers et al., 2009; Stratton et al., 2005). Medical specialty selection is the biggest and most enduring decision made during their tenure at medical school, and is on par with choosing medicine as a career in general (Borges, 2007; Reed et al., 2001). To prepare them to eventually work in their chosen medical specialty, medical students receive similar core classes while in medical school. Between 60% to 75% of medical students change their specialty while still in medical school (Savickas et al., 1986; Markert, 1983) because they lack sufficient experience and information to make the decision (Savickas et al., 1988). Even after the residency choice is made, changes are still likely to occur. Changing a residency not only costs hospitals time and money in training the resident, but the resident becomes frustrated and loses time and money while finding and then completing a second residency (Borges et al., 2005).

For this study, the Medical Specialty Preference Inventory-Revised (MSPI-R) (Richard, 2011) was utilized to compare validity, including hit rates, by gender for its current scoring system, based upon the occupational groups of Strong, to Kuder's person matching model. Validity implications for each gender are required before considerable resources will be expended to study the full person matching protocol including career narratives (including current occupation, past occupations, lifestyle, future goals, typical day, needed skills, and descriptions of what they like best and least about their occupation) with currently established career interest inventories. Further, research has yet to examine the MSPI-R for gender differences in validity. Additionally, person matching could deliver results for every medical specialty instead of only the 16 offered today with the MSPI-R. This would allow medical students to have the possibility to be person matched to the full range of medical specialties to help foster medical specialty exploration, which has been called for in the literature (Borges & Savickas, 2002).

3. Method

3.1. Participants

The participants were 5143 (2898 female and 2245 male) medical students enrolled in medical schools across the United States who took the MSPI-R between January 2005 and April 2008 from several ethnicities; White (3447), Asian (767), African American (343), Hispanic (250), Other (53), American Indian/Alaska Native (27), and Native Hawaiian/Pacific Islander (6) with 250 participants not identifying an ethnicity. Ages ranged between 57 and 24 at the time of taking the MSPI-R with an average age of 30 ($SD = 3.22$). The medical students eventually practiced in 44 medical specialties ranging from Internal Medicine (1007 medical students) to Medical Toxicology (1 medical student) with an average of 117 medical students per medical specialty ($SD = 213$) as listed in **Table 1**.

3.2. Random Sample

To test both psychometric scoring methodologies to be inclusive of the largest number of medical specialties (the literature calls for including more than 16 medical specialties), a subset of the full reference group had to be selected and then compared individually to the entire reference group. By allowing all members of the reference group to be part of the random sample selection process, the researcher mimicked the reality that not all medical students taking the MSPI-R enter the medical specialty calculated by the inventory. A stratified random sample of 500 medical students (250 females and 250 males) was chosen from the reference group of 5143 medical students to ensure a high confidence level and low confidence interval. Twenty-two medical specialties were selected to be part of the stratified random sample as they contained at least 10 medical students (five male and female) who had entered the medical specialty. Since person matching requires that occupations must be represented in the reference group for a match to be made, one person was removed from the entire reference group at a time so that at least nine other individuals from a medical specialty were represented in the reference group. It would significantly hinder the hit rate of person matching to randomly remove 500 people from the reference group when seven of them may have come from a specialty represented by 10 individuals. To further mimic real world processes in the study, the researcher matched the proportion of medical students in each

Table 1. Frequencies of medical residency selections made by the reference group (N = 5143) and the criterion group (N = 500).

Residency choice	Reference group	Reference group	Criterion group	Criterion group
Internal medicine	1007	19.58	56	11.2
Pediatrics	695	13.51	44	8.8
Emergency medicine	514	9.99	40	8.0
Family medicine	507	9.86	40	8.0
Obstetrics/gynecology	379	7.37	30	6.0
Surgery	330	6.42	30	6.0
Anesthesiology	297	5.77	24	4.8
Psychiatry	245	4.76	24	4.8
Orthopedic surgery	201	3.91	24	4.8
Radiology	194	3.77	20	4.0
Pathology	149	2.90	20	4.0
Internal medicine pediatrics	139	2.70	20	4.0
Otolaryngology	94	1.83	14	2.8
Ophthalmology	65	1.26	14	2.8
Neurology	64	1.24	14	2.8
Dermatology	58	1.13	14	2.8
Physical medicine and rehabilitation	55	1.07	14	2.8
Urology	33	0.64	14	2.8
Neurological surgery	27	0.52	14	2.8
Radiation oncology	18	0.35	10	2.0
Plastic surgery	16	0.31	10	2.0
Pediatrics/psychiatry/child and adolescent psychiatry	11	0.21	10	2.0
Internal medicine/psychiatry	4	0.08	0	0.0
Preventive medicine	4	0.08	0	0.0
Psychiatry family practice	4	0.08	0	0.0
Pediatrics/physical medicine and rehabilitation	3	0.06	0	0.0
Child and adolescent psychiatry	2	0.04	0	0.0
Infectious disease	2	0.04	0	0.0
Child neurology	2	0.04	0	0.0
Internal medicine/dermatology	2	0.04	0	0.0
Internal medicine/emergency medicine	2	0.04	0	0.0
Internal medicine/family practice	2	0.04	0	0.0
Internal medicine/neurology	2	0.04	0	0.0
Nuclear medicine	2	0.04	0	0.0

Continued

Pediatric emergency medicine	2	0.04	0	0.0
Pediatric hematology/oncology	2	0.04	0	0.0
Vascular surgery	2	0.04	0	0.0
Pediatrics/medical genetics	2	0.04	0	0.0
Cardiovascular disease	1	0.02	0	0.0
Gastroenterology	1	0.02	0	0.0
Medical toxicology	1	0.02	0	0.0
Pediatrics/dermatology	1	0.02	0	0.0
Psychiatry/neurology	1	0.02	0	0.0
Sports medicine	1	0.02	0	0.0
Total	5143	100	500	100

medical specialty in the reference group to the proportion of medical students in the 22 medical specialties chosen to be part of the stratified random sample.

3.3. Measure

The Medical Specialty Preference Inventory-Revised (MSPI-R) measures interest in 18 areas of medical practice and predicts entrance into 16 major medical specialties. The MSPI-R provides information to medical students to help them choose a medical specialty appropriate to their interests following graduation from medical school. To take the MSPI-R, a medical student selects one of seven scale points to indicate the degree of desirability for each item on the inventory; the next item is displayed until the MSPI-R is completed. Medical students instantaneously receive a report of results including 16 Medical Specialty Choice Probabilities along with 18 Medical Interest Scales (Richard, 2011). For each of the 16 medical specialties, a percentage score is reported that indicates the likelihood that the student will enter into the specialty. The 16 medical specialty percentages, when added together, total 100 percent and are presented in order from the highest to the lowest likelihood that the student would enter each specialty (Richard, 2011). Students are instructed to select the two or three specialties with the highest probabilities to explore further. Next, students receive their Medical Interest Scale scores to identify their highest and lowest scoring interests in 18 areas of medical practice that are experienced in varying degrees in each medical specialty.

There are 150 items included in the MSPI-R; however, only 102 items are used to score the instrument (Richard, 2011). Of those, 88 items are used to score the 18 Medical Interest Scales, and 30 items are used to score the 16 Specialty Choice Probabilities (Richard, 2011). Sixteen of the items are scored in both the Medical Interest Scales and the Specialty Choice Probabilities (Richard, 2011). The remaining 48 items are not scored, and may be used in the future for possible replacement of items as needed to improve the ability of the instrument and to support the development of new specialties (Richard, 2011). The 18 Medical Interest Scales include Complex Problems, Comprehensive Care, Diagnostic Precision, Emergency-Critical Care, History Taking, Home Health Care, Immediate Results, Knowledge of Anatomical Structures, Knowledge of Organ Systems, Laboratory Results, Palliative Care, Patient Counseling, Prevention and Education, Procedural Care, Psychological Care, Reproductive Care, Social Context, and Technology in Medicine. In addition, the MSPI-R calculates preferences for 16 medical specialties: Anesthesiology, Dermatology, Emergency Medicine, Family Medicine, Internal Medicine, Neurology, Obstetrics and Gynecology, Orthopedic Surgery, Otolaryngology, Pathology, Pediatrics, Physical Medicine and Rehabilitation, Psychiatry, Radiology, Surgery, and Urology.

Cronbach's alpha, a measure of internal consistency, suggested reliabilities ranging from a low of 0.77 for History Taking and Diagnostic Precision to 0.94 for Psychological Care (Richard, 2011). Comparisons between the 2nd edition MSPI factors and the revised MSPI-R Medical Interest Scales suggested high positive correlations and indicated sufficient validity of the new MSPI-R Medical Interest Scales (Richard, 2011).

3.4. Data Analysis Procedures

The MSPI-R's raw scores were analyzed by gender in two different ways: a) person matching using all 150 items on the inventory and b) the standard method of scoring the MPSI-R. All person matching analyses used Cronbach and Gleser's (1953) difference squared (D2) values, a person matching statistic, to determine the linear distance of profile similarity. D2, the sum of squared Euclidian distances between self- and other-ratings of traits, reflects differences in elevation, scatter, and shape between individuals' scores on the same inventory (Cronbach & Gleser, 1953). The D2 statistic is a descriptive statistic and as such does not include the concepts of statistical power or statistical precision. When using the D2 statistic, score 1 (from a test taker) is subtracted from score 2 (an individual from a reference group) with the resulting difference being squared (Hartung et al., 2005). When the differences between the two scores are squared, the result becomes normally distributed (Cronbach & Gleser, 1953). In the D2 statistic there is no upper limit on the distance between two scores. However, the closer the D2 calculation is to 0, the closer the two individuals scored similarly on the inventory, which signifies a close person-to-person match.

D2 values were calculated by subtracting item scores for each member of the stratified random sample from the item scores of each individual in the reference group with the resulting differences being squared. The squared differences comparing the two individuals were summed to obtain a final score. Scores comparing the test taker to each of the 5142 members of the reference group were then placed into rank order from the lowest to the highest.

To record the calculations for person matching, hit rates by gender were recorded based upon a rank ordered list comparing each member of the stratified random sample to the 5142 members of the reference group and documenting the closest five matches. To record the calculations for standard scoring, the five highest specialty choice probabilities were generated by gender using beta weights with 30 out of the total 150 MSPI-R items. Third, coefficient kappa was calculated to determine the interrater agreement between the predicted and the actual medical specialty selection beyond chance for the first match for each of the two scoring methods by gender. Lastly, chance expectancy hit rates were calculated to determine the expected hit rates by gender that would be achieved by chance alone for both psychometric scoring methodologies.

4. Results

Table 2 displays hit rates by gender for person matching and standard scoring based on the five top predictions for the 500 members of the random sample. Data suggest that when calculated using person matching, female and male hit rates were comparable for the first, second, and fifth matches. Females received 2% fewer hits for the third match and 1% fewer hits for the fourth match with person matching. Data suggest that when calculated using standard scoring, females received 6% fewer first match hits as compared to males. Females received 1% additional hits for the second match, 3% additional hits for the third and fifth matches, and comparable hits for the fourth match with standard scoring. Hypothesis 1 suggested that hit rates by gender for person matching would be equal between genders and unequal between genders for standard scoring when using the Medical Specialty Preference Inventory-Revised with medical students. No significant difference was found between genders and hit rates. Hypothesis 2 suggested hit rates for females would be higher for person matching as compared to female hit rates for standard scoring when using the Medical Specialty Preference Inventory-Revised with medical students. No significant difference was found between females and the hit rates of the two scoring methodologies.

Kappa coefficients for first match hit rate accuracy between actual versus predicted medical specialty for standard scoring and person matching by gender were calculated. A kappa value of less than 0.20 represents poor agreement; between 0.21 and 0.40 represents fair agreement; between 0.41 and 0.60 represents moderate agreement; between 0.61 and 0.80 represents good agreement; and between 0.81 and 1.0 represents very good agreement beyond chance (Landis & Koch, 1977). The interrater reliability for standard scoring for females was $\kappa = 0.28$ ($p < 0.001$), 95% CI (0.21, 0.35). The interrater reliability for standard scoring for males was $\kappa = 0.37$ ($p < 0.001$), 95% CI (0.30, 0.44). The interrater reliability for person matching for females suggested $\kappa = 0.19$ ($p < 0.001$), 95% CI (0.13, 0.25). The interrater reliability for person matching for males suggested $\kappa = 0.18$ ($p < 0.001$), 95% CI (0.12, 0.24). Standard scoring for males obtained the highest kappa coefficient values. Further, higher kappa coefficient values were achieved by standard scoring for males and females as compared to person matching. However, person matching kappa coefficient values were more consistent between

Table 2. Hit rates by gender.

Criteria		<i>n</i>	Hit rate
Female	First	56/250	22%
	Second	23/250	10%
	Third	16/250	6%
	Fourth	12/250	5%
	Fifth	16/250	6%
	Total of top 5	123/250	49%
Male	First	55/250	22%
	Second	24/250	10%
	Third	20/250	8%
	Fourth	14/250	6%
	Fifth	15/250	6%
	Total of top 5	128/250	51%
Standard scoring			
Females	First	74/250	30%
	Second	34/250	14%
	Third	24/250	10%
	Fourth	18/250	7%
	Fifth	17/250	7%
	Total of top 5	167/250	67%
Males	First	91/250	36%
	Second	33/250	13%
	Third	17/250	7%
	Fourth	18/250	7%
	Fifth	9/250	4%
	Total of top 5	168/250	67%

males and females. Hypothesis 3 stated that kappa coefficient values would be equal between genders for person matching and unequal between genders for standard scoring when using the Medical Specialty Preference Inventory-Revised with medical students. This was not supported as no significant differences were detected. Hypothesis 4 stated that kappa coefficient values for females would be higher for person matching as compared to female hit rates for standard scoring when using the Medical Specialty Preference Inventory-Revised with medical students. This was unsupported as kappa coefficient values for females were higher for standard scoring.

Table 3 displays the expected hit rates by gender that would be achieved by chance for the first match with standard scoring and person matching. Person matching and standard scoring accurately placed male and female medical students at a rate greater than chance into 15 out of 22 medical specialties. Standard scoring is limited to placing male and female medical students at a rate greater than chance into only 16 medical specialties as predictions to occupations can only be made for 16 medical specialties. However, while standard scoring can make predictions for Neurology as well as Physical Medicine and Rehabilitation, it fell below the chance expectancy rate for Neurology for males and Physical Medicine and Rehabilitation for females. Kuder stated that an advantage of person matching was the ability to place an unlimited amount of occupations in the reference group im-

Table 3. Chance expectancy hit rates for the first match by specialty and gender.

Medical specialty <i>n</i> chance standard person matching					
Females					
	Internal medicine	28/250	0.11	0.47	0.32
	Pediatrics	22/250	0.09	0.50	0.32
	Emergency medicine	20/250	0.08	0.40	0.25
	Family medicine	20/250	0.08	0.55	0.40
	Obstetrics/gynecology	15/250	0.06	0.53	0.40
	Surgery	15/250	0.06	0.33	0.27
	Anesthesiology	12/250	0.05	0.08	0.08
	Psychiatry	12/250	0.05	0.17	0.08
	Orthopedic surgery	12/250	0.05	0.25	0.50
	Radiology	10/250	0.04	0.05	0.20
	Pathology	10/250	0.04	0.20	0.20
	Internal medicine pediatrics	10/250	0.04	0.00	0.10
	Otolaryngology	7/250	0.03	0.14	0.14
	Ophthalmology	7/250	0.03	0.00	0.00
	Neurology	7/250	0.03	0.14	0.14
	Dermatology	7/250	0.03	0.14	0.29
	Physical medicine & rehab	7/250	0.03	0.14	0.00
	Urology	7/250	0.03	0.00	0.00
	Neurological surgery	7/250	0.03	0.00	0.00
	Radiation oncology	5/250	0.02	0.00	0.00
	Plastic surgery	5/250	0.02	0.00	0.00
	Peds/child & adolescent psy	5/250	0.02	0.00	0.00
Number of specialties below chance expectancy rate		7	7		
Males					
	Internal medicine	28/250	0.11	0.45	0.29
	Pediatrics	22/250	0.09	0.68	0.27
	Emergency medicine	20/250	0.08	0.45	0.50
	Family medicine	20/250	0.08	0.50	0.30
	Obstetrics/gynecology	15/250	0.06	0.27	0.13
	Surgery	15/250	0.06	0.60	0.27
	Anesthesiology	12/250	0.05	0.42	0.08
	Psychiatry	12/250	0.05	0.67	0.42
	Orthopedic surgery	12/250	0.05	0.33	0.42

Continued

Radiology	10/250	0.04	0.10	0.10
Pathology	10/250	0.04	0.20	0.20
Internal medicine pediatrics	10/250	0.04	0.00	0.00
Otolaryngology	7/250	0.03	0.43	0.43
Ophthalmology	7/250	0.03	0.00	0.00
Neurology	7/250	0.03	0.00	0.14
Dermatology	7/250	0.03	0.14	0.14
Physical medicine & rehab	7/250	0.03	0.06	0.00
Urology	7/250	0.03	0.14	0.14
Neurological surgery	7/250	0.03	0.00	0.00
Radiation oncology	5/250	0.02	0.00	0.00
Plastic surgery	5/250	0.02	0.00	0.00
Peds/child & adolescent psy	5/250	0.02	0.00	0.00
Number of specialties below chance expectancy rate			7	7

mediately to allow for predictions in occupations, which would signal an improvement in besting chance expectancy rates. It is surprising that person matching equaled standard scoring in the ability to best chance expectancy rates. This could be due to the limits of the way person matching was performed in this study, which are fully articulated later in recommendations for improvements.

5. Discussion

A previous study of standard scoring with the MSPI-R suggested a first match hit rate of 54% (Porfeli et al., 2010). The MSPI-R manual suggests a first match hit rate of 52% (Richard, 2011). These reported hit rates for the MSPI-R are higher than career interest inventories predicting general occupational choice, such as the Self Directed Search with a hit rate of 47% (Glavin & Savickas, 2011) and the 2005 Strong Interest Inventory with a hit rate of 38% (Gasser et al., 2007). For standard scoring with the MSPI-R in this study, results suggested that males achieved a first match hit rate of 36% and females achieved a first match hit rate of 30%. It is unknown why this study's standard scoring hit rates for the MSPI-R align more closely with the hit rates of career interest inventories predicting general occupational choice than the previous two cited sources for the MSPI-R. For person matching in this study, results suggested that males and females achieved a first match hit rate of 22%. Using the current calculation strategies for both methodologies, standard scoring clearly yields higher hit rates for both genders.

The data proposes that when using standard scoring, females receive 6% lower first match hits as compared to males with the MSPI-R in the US, however, these differences are not significant. As the standard scoring method used with the MSPI-R does not have separate scoring norms for males and females, the possibility is raised that standard scoring with the MSPI-R should be tested to incorporate separate gendered scoring norms to see if hit rates improve for females. However, when looking at predictive accuracy by the fifth match, both standard scoring and person matching appear gender neutral. Additionally, both psychometric scoring methodologies struggle with first match hit rate accuracy by gender with the MSPI-R and demonstrate lower kappa coefficient accuracy than the reported hit rates. Work to improve the psychometric scoring methodology of the MSPI-R for females and males needs to be continued.

Although the hit rate accuracy for person matching with females and males in this study was lower than the hit rate of standard scoring, validity increased using person matching as more medical specialties had the potential for use in the career interest inventory. Additionally, new occupations can be added immediately as a hand-

ful of individuals work in the occupation with person matching. This is in sharp contrast to standard scoring requiring several hundred individuals to work in an occupation for an occupational scale to be developed. The increase in narrative data received from the person matching scoring report (including current occupation, past occupations, lifestyle, future goals, typical day, needed skills, and descriptions of what they like best and least about their occupation) could help males and females construct their preferred career story by providing a narrative context from which to make meaning, clarify values, and solidify decisions.

The Association of American Medical Colleges (AAMC) would likely profit by providing students with person matching as part of the MSPI-R scoring report to further assist medical specialty decision making for female and male medical students. Person matching offers the greatest ability to assign medical students to membership in an occupational group because all 44 medical specialties in the sample could be predicted with the MSPI-R while standard scoring could only make predictions for 16 medical specialties. Hence, person matching increases career interest inventory validity over standard scoring as it has the greater ability to a) differentiate between and b) assign to specific occupational groups for females and males. While the data suggest more improvements need to be made to the psychometric scoring methodology, person matching in career interest inventories deserves research attention to benefit women and the career decision making needs of a quickly changing, global workforce.

6. Limitations

There are several limitations to this research. First, this study has only compared validity by gender between standard scoring and person matching. Person matching could demonstrate the full range of its advantages over the scoring report of standard scoring by performing a study for each gender comparing their experiences with the two score reports.

The reference group contained an imbalanced number of individuals in each medical specialty. Internal Medicine had the highest number with 1007 members and there were six medical specialties with one member. The extreme range of medical students in each medical specialty in the reference group may have created disproportions with person matching and impacted hit rate accuracy by gender.

Third, Kuder limited the reference group for person matching to individuals who were enthusiastic about their work so that test takers would be matched to passionate people who scored the career interest inventory similarly (Kuder, 1977a). This research did not determine if medical students in the reference group were passionately working in their medical specialty, which may have hindered person matching's hit rate performance.

Fourth, medical students who voluntarily took the MSPI-R between the years 2005 and 2008 comprise the sample. The sample is not representative of all medical students and may be made up of a specific type of medical student as medical students are not required to take the MSPI-R. Medical students at medical colleges that are part of the Association of American Medical Colleges (AAMC) were able to access the MSPI-R and participate in the study. Accordingly, there is no ability to generalize results to medical students attending medical schools outside of the United States.

7. Recommendations for Future Research

In addition to comparing female and male test takers' views on the value of the two different scoring reports produced by standard scoring and person matching, additional questions have been uncovered by this research. Comparing person matching validity by gender for all individuals in the reference group versus only individuals enthusiastic about their work in the reference group would help to verify Kuder's assertion about enthusiasm and hit rate accuracy. Additionally, an investigation is needed into even versus uneven reference group occupational membership to determine changes in hit rate accuracy by gender for person matching. Lastly, this researcher did not examine how standard scoring and person matching hit rates compared for medical students of different ethnicities and cultures.

Acknowledgements

Special thanks to the Association of American Medical Colleges for permitting the use of their data to perform this study along with Mark L. Savickas, George V. Richard, and Erik J. Porfeli for their support during this process.

References

- Amundson, N. E. (2009). *Active Engagement: The Being and Doing of Career Counselling* (3rd ed.). Richmond: Ergon Communications.
- Armstrong, P. I., & Rounds, J. (2010). Integrating Individual Differences in Career Assessment: The Atlas Model of Individual Differences and the Strong Ring. *Career Development Quarterly*, *59*, 143-153. <http://dx.doi.org/10.1002/j.2161-0045.2010.tb00058.x>
- Borges, N. (2007). Behavioral Exploration of Career and Specialty Choice in Medical Students. *Career Development Quarterly*, *55*, 351-358. <http://dx.doi.org/10.1002/j.2161-0045.2007.tb00089.x>
- Borges, N., & Savickas, M. L. (2002). Personality and Medical Specialty Choice: A Literature Review and Integration. *Journal of Career Assessment*, *10*, 362-380. <http://dx.doi.org/10.1177/10672702010003006>
- Borges, N., Gibson, D., & Karnani, R. (2005). Job Satisfaction of Physicians with Congruent versus Incongruent Specialty Choice. *Evaluation and the Health Professions*, *28*, 400-413. <http://dx.doi.org/10.1177/0163278705281071>
- Campbell, D., & Borgen, F. (1999). Holland's Theory and the Development of Interest Inventories. *Journal of Vocational Behavior*, *55*, 86-101. <http://dx.doi.org/10.1006/jvbe.1999.1699>
- Cochran, L. (1997). *Career Counseling: A Narrative Approach*. Thousand Oaks, CA: Sage.
- Cronbach, L., & Gleser, G. (1953). Assessing Similarity between Profiles. *The Psychological Bulletin*, *50*, 456-473. <http://dx.doi.org/10.1037/h0057173>
- Donnay, D. (1997). E.K. Strong's Legacy and beyond: 70 Years of the Strong Interest Inventory. *Career Development Quarterly*, *46*, 2-22. <http://dx.doi.org/10.1002/j.2161-0045.1997.tb00688.x>
- Einarsdóttir, S., & Rounds, J. (2009). Gender Bias and Construct Validity in Vocational Interest Measurement: Differential Item Functioning in the Strong Interest Inventory. *Journal of Vocational Behavior*, *74*, 295-307. <http://dx.doi.org/10.1016/j.jvb.2009.01.003>
- Gasser, C. E., Larson, L. M., & Borgen, F. H. (2007). Concurrent Validity of the 2005 Strong Interest Inventory: An Examination of Gender and Major Field of Study. *Journal of Career Assessment*, *15*, 23-43. <http://dx.doi.org/10.1177/1069072706294516>
- Glavin, K., & Savickas, M. L. (2011). Interpreting Self-Directed Search Profiles: Validity of the Rule of Eight. *Journal of Vocational Behavior*, *79*, 414-418. <http://dx.doi.org/10.1016/j.jvb.2011.01.005>
- Hackett, G., & Lonborg, S. D. (1993). Career Assessment for Women: Trends and Issues. *Journal of Career Assessment*, *1*, 197-216. <http://dx.doi.org/10.1177/106907279300100301>
- Harmon, L. W. (1997). Do Gender Differences Necessitate Separate Career Development Theories and Measures? *Journal of Career Assessment*, *5*, 463-470. <http://dx.doi.org/10.1177/106907279700500406>
- Hartung, P., Borges, N., & Jones, B. (2005). Using Person Matching to Predict Career Specialty Choice. *Journal of Vocational Behavior*, *67*, 102-117. <http://dx.doi.org/10.1016/j.jvb.2004.03.001>
- Holland, J. L. (1961). Some Explorations with Occupational Titles. *Journal of Counseling Psychology*, *8*, 82-87. <http://dx.doi.org/10.1037/h0042428>
- Holland, J. L. (1966). *The Psychology of Vocational Choice: A Theory of Personality Types and Model Environments*. Waltham, MA: Blaisdell Publishing Company.
- Holland, J. L., Whitney, D., Cole, N., & Richards, J. J. (1969). *An Empirical Occupational Classification Derived from a Theory of Personality and Intended for Practice and Research*. Iowa City, IA: American College Testing Program.
- Hyde, J. S. (2005). The Gender Similarities Hypothesis. *American Psychologist*, *60*, 581-592. <http://dx.doi.org/10.1037/0003-066X.60.6.581>
- Ihle-Helledy, K., Zytowski, D., & Fouada, N. (2004). Kuder Career Search: Test-Retest Reliability and Consequential Validity. *Journal of Career Assessment*, *12*, 285-297. <http://dx.doi.org/10.1177/1069072703257752>
- Kuder, F. (1977a). *Activity Interests and Occupational Choice*. Chicago, IL: Science Research Associates.
- Kuder, F. (1977b). Career Matching. *Personnel Psychology*, *30*, 1-4. <http://dx.doi.org/10.1111/j.1744-6570.1977.tb02315.x>
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*, 159-174. <http://dx.doi.org/10.2307/2529310>
- Li, C., & Kerpelman, J. (2007). Parental Influences on Young Women's Certainty about Their Career Aspirations. *Sex Roles*, *56*, 105-115. <http://dx.doi.org/10.1007/s11199-006-9151-7>
- Lippa, R. (1998). Gender-Related Individual Differences and the Structure of Vocational Interests: The Importance of the People—Things Dimension. *Journal of Personality and Social Psychology*, *74*, 996-1009. <http://dx.doi.org/10.1037/0022-3514.74.4.996>

- Markert, R. J. (1983). Stability and Change of Medical Specialty Choice in U.S. Medical Schools. *Journal of Medical Education*, 58, 589-590.
- McMahon, M., & Watson, M. (2010). Story Telling: Moving from Thin Stories to Thick and Rich Stories. In K. Maree (Ed.), *Career Counselling: Methods That Work* (pp. 53-63). Cape Town: Juta.
- Nauta, M. M. (2010). The Development, Evolution, and Status of Holland's Theory of Vocational Personalities: Reflections and Future Directions for Counseling Psychology. *Journal of Counseling Psychology*, 57, 11-22. <http://dx.doi.org/10.1037/a0018213>
- Peavy, R. V. (1998). *SocioDynamic Counselling: A Constructivist Perspective*. Victoria: Trafford.
- Porfeli, E. J., Richard, G. V., & Savickas, M. L. (2010). Development of Specialization Scales for the MSPI: A Comparison of Empirical and Inductive Strategies. *Journal of Vocational Behavior*, 77, 227-237. <http://dx.doi.org/10.1016/j.jvb.2010.04.011>
- Reed, V. A., Jernstedt, G. C., & Reber, E. S. (2001). Understanding and Improving Medical Student Specialty Choice: A Synthesis of the Literature Using Decision Theory as a Referent. *Teaching and Learning in Medicine*, 13, 117-129. http://dx.doi.org/10.1207/S15328015TLM1302_7
- Richard, G. (2011). *Medical Specialty Preference Inventory, Revised Edition Technical Manual*. Washington DC: Association of American Medical Colleges.
- Rogers, M. E., Creed, P. A., & Searle, J. (2009). The Development and Initial Validation of Social Cognitive Career Theory Instruments to Measure Choice of Medical Specialty and Practice Location. *Journal of Career Assessment*, 17, 324-337. <http://dx.doi.org/10.1177/1069072708330676>
- Savickas, M. L., Alexander, D., Jonas, A., & Wolf, F. (1986). Difficulties Experienced by Medical Students in Choosing a Specialty. *Journal of Medical Education*, 61, 467-469.
- Savickas, M. L., Brizzi, J., Brisbin, L., & Pethel, L. (1988). Predictive Validity of Two Medical Specialty Preference Inventories. *Measurement and Evaluation in Counseling and Development*, 21, 106-112.
- Savickas, M. L., Nota, L., Rossier, J., Dauwalder, J., Duarte, M. E., Guichard, J., Soresi, S., Van Esbroeck, R., & van Vianen, A. E. M. (2009). Life Designing: A Paradigm for Career Construction in the 21st Century. *Journal of Vocational Behavior*, 75, 239-250. <http://dx.doi.org/10.1016/j.jvb.2009.04.004>
- Sodano, S., & Richard, G. (2009). Construct Validity of the Medical Specialty Preference Inventory: A Critical Analysis. *Journal of Vocational Behavior*, 74, 30-37. <http://dx.doi.org/10.1016/j.jvb.2008.10.001>
- Stockdale, M. S., & Nadler, J. T. (2012). Paradigmatic Assumptions of Disciplinary Research on Gender Disparities: The Case of Occupational Sex Segregation. *Sex Roles*, 68, 207-215. <http://dx.doi.org/10.1007/s11199-012-0228-1>
- Stratton, T. D., Witzke, D. B., Elam, C. L., & Cheever, T. R. (2005). Learning and Career Specialty Preferences of Medical School Applicants. *Journal of Vocational Behavior*, 67, 35-50. <http://dx.doi.org/10.1016/j.jvb.2003.06.001>
- Strong, E. K. (1943). *Vocational Interests of Men and Women*. Stanford, CA: Stanford University Press.
- Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and Things, Women and People: A Meta-Analysis of Sex Differences in Interests. *Psychological Bulletin*, 135, 859-884. <http://dx.doi.org/10.1037/a0017364>
- Wetzel, E., Hell, B., & Passler, K. (2012). Comparison of Different Test Construction Strategies in the Development of a Gender Fair Interest Inventory Using Verbs. *Journal of Career Assessment*, 20, 88-104. <http://dx.doi.org/10.1177/1069072711417166>
- Whiston, S. C., & Bouwkamp, J. C. (2003). Ethical Implications of Career Assessment with Women. *Journal of Career Assessment*, 11, 59-75. <http://dx.doi.org/10.1177/106907202237460>