2012

Cultural Biases in the Weschler Memory Scale iii (WMS-iii)

Adam David Less
University of North Florida

CULTURAL BIASES IN THE WESCHLER MEMORY SCALE iii (WMS-iii)

by

Adam David Less

B.S., University of North Florida, 2008

A thesis submitted to the Department of Psychology
in partial fulfillment of the requirements for the degree of

Master of Arts in General Psychology

UNIVERSITY OF NORTH FLORIDA

COLLEGE OF ARTS AND SCIENCES

February 24, 2012

Certificate of Approval

The thesis of Adam David Less is approved: (Date)

Signature Deleted

4/2/12

Thesis Advisor (Dr. Michael Toglia)

Signature Deleted

4/3/12

2nd Reader (Dr. Randy Russac)

Accepted For The Department:

Signature Deleted

4/2/12

Chair (Dr. Michael Toglia)

Accepted For The College:

Signature Deleted

9 April 2012

Dean

Accepted For The University:

Signature Deleted

5/2/12

Dean of the Graduate School

Acknowledgments

I would like to thank all of the faculty, staff and my cohorts at the University of North Florida. Without the education and experiences I gained from these relationships, I may have never acquired this level of understanding and achievement in the academic world. Additionally, I would like to make special mentions of Dr. Michael Toglia, my primary thesis advisor and LeAnn Anderson, M.A., whom is both a colleague and friend. Dr. Toglia became my thesis advisor through a series of unforeseen events, and has always maintained a level of patience and professionalism that is wanting by professionals in so many other fields of study. Ms. Anderson became acquainted with me in the early years of my graduate career, and has played a major role in the development of my analytical abilities and the completion of this thesis. Without the advice of Ms. Anderson or Dr. Toglia, this thesis may not have been completed in the high quality fashion that readers will discover. I would also like to thank my family for their unwavering support in all my academic ventures.

Table of Contents

List of Tables

List of Figures

Abstract

The Wechsler Memory Scale –iii is the newest version of a six-decade old neuropsychological inventory. Since its conception, the Wechsler Memory Scale has been highly utilized by practitioners to accurately assess various memory functions in adult subjects. Revisions made within this inventory include the Faces I subtest, a facial recognition scale, which was added in order to strengthen the instrument's accuracy at measuring episodic memory. Facial recognition, both cross-race and within-race, has been researched extensively and consistent biases have been found between race of test taker and cross-racial identification. Theories of exposure/contextual interaction (environment) and biological foundations have been the subject of study in the past in order to determine from where these racial identification deficits stem. The current study focuses on revealing bias in the Faces I subtest, regarding to an unequal distribution of racially representative faces in the testing materials. Eighty-eight college students were recruited to view forty-eight pictured faces from the Faces I subtest and determine the racial category to which the pictured face belonged. The subjects' categorical responses were the basis for calculating a percent agreement score for racial category of each face. It was determined, using the results of subjects' responses, that the Faces I subtest contained an unequal distribution of racially representative faces in both the Target and Interference testing material. This confirmed the presence of an inherent bias within the subscale. The implications of memory accuracy for the WMS-iii are discussed as it relates to different fields of study, but none more directly than the criminal justice system. Eyewitness testimony is a pivotal evidentiary tool in the criminal justice system, and ramifications of cross-racial identification deficits and biases in the tools to accurately assess memory are increasingly bringing this once heavily relied upon tool into question.

Keywords: Wechsler memory scale-iii, cross-racial identification, eyewitness testimony, facial recognition

Cultural Biases in the Wechsler Memory Scale-iii (WMS-iii)

In the area of neuropsychology, psychologists employ numerous auditory and visual tests to assess mental function. Wechsler inventories have been used to assess various mental functions for decades (Wechsler, 1997). Each Wechsler test inventory has a specific purpose and application. The Wechsler Adult Intelligence Scale-iii (WAIS-iii), for example, is employed to measure adult and adolescent intelligence (Wechsler, 1997). Another widely used test battery for mental assessment is the Wechsler Memory Scale (WMS; Wechsler, 1997). Appropriately, the Wechsler Memory Scale (WMS) has been employed since the 1940's to measure different components of memory functioning in adolescents and adults (Opasanon, 2008). The original WMS inventory received criticism regarding the adequacy of the standardization sample (n=200), validity of scores obtained and the overall reliability of the instrument (Franzen, 2000). In order to improve the scale, revisions have periodically been made to the overall instrument and its scoring system (Franzen, 2000). The most recently updated version of the memory test, the WMS-iii, is primarily used in providing a more detailed analysis of declarative memory function (Wechsler, 1997). Declarative memory, also known as explicit memory, refers to memories which can be consciously recalled such as facts and events (Ullman, 2004). This type of memory function is increasingly being studied, as deficiencies in declarative memory are seen in various debilitating diseases, such as Alzheimer’s disease. On a side note, when used in conjunction with the WAIS-iii (Wechsler, 1997), the WMS-iii can yield meaningful comparisons between intellectual ability and memory function (Wechsler, 2004).

The WMS-iii consists of four co-normed episodic memory tests (i.e., Logical Memory, Faces, Verbal Paired Associates, and Family Pictures) that yield eight age- and demographically-adjusted standard scores (Noor & Najam, 2009). Episodic memory is unique, as it captures information such as “what”, “when” and “where” (Noor & Najam, 2009). It is the memory structure that remembers stimuli that is observed through experience, and each test is used in the Faces I subtest is meant to assess one or more of the following eight primary memory indices: auditory immediate, visual immediate, immediate memory, auditory delayed, visual delayed, auditory recognition delayed, general memory and working memory (Noor & Najam, 2009). The Faces I subtest, a new addition to the Wechsler Memory Scale, conforms to a recognition paradigm allowing practitioners to measure immediate and delayed memory by administering a visual-recognition assessment and evaluating recognition deficits (Wechsler, 2004). This type of evaluation permits researchers to assess immediate recall of a subject’s memory as well as decay in delayed recognition when the assessment is re-administered.

A major advantage to adding facial memory scales to neuropsychological inventories is that they may assist in rehabilitation services due to the high levels of ecological and face validity (O’Bryant & McCaffrey, 2006). Analogous to real world settings, facial memory scales provide a simultaneous measure of multiple memory functions. Instead of isolating working memory, facial memory scales allow a researcher and/or clinician to monitor visual immediate and delayed memory as well as general immediate memory (O’Bryant & McCaffrey, 2006; Wechsler, 1997).

As with any standardized test, the WMS-iii must be administered within the exact parameters of its design. Administration of the Faces I subtest begins with the examiner exposing a series of 24 target faces, one at a time, for 2 seconds each and asking the patient or subject to remember each face. A second series of 48 faces (24 interference faces not previously presented and the original 24 target faces) is then shown to the subject. The subject is told to identify each face as either one they were asked to remember (target face) or one that is a new (interference) face , and the test administrator scores the subject’s answers on a “hit or miss” scale (Wechsler, 1997).

There has been surprisingly a paucity of literature investigating the psychometric properties of the WMS-iii inventory. Some researchers have criticized the Faces I subtest because it is uncorrelated with other WMS-iii visual memory assessments and that it fails to differentiate between clinical groups, i.e., dementia patients, elderly patients, groups with mental disease or defect (e. g., Holdnack & Delis, 2004; Migoya, Zimmerman & Golden, 2002; Wechsler, 1997). Holdnack and Delis (2004) examined these criticisms by implementing four individual studies evaluating the utility of applying signal detection measures to the face memory subtest (Parra, Abrahams, Fabi, Logie, Luzzi, & Della Sala, 2009; Parra, Abrahams, Logie, Mendez, Lopera, & Della Sala, 2010; Wixted, 2007; Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996). The first two studies involved the WMS-iii standardization data set to determine age and education effects and to present normative and reliability data for hits, false positives, discriminability and response bias (Holdnack & Delis, 2004). The third study was designed to test the hypothesis that using response components, such as “Yes” and “No”, and signal detection measures (measure derived from signal detection theory, wherein accuracy can be determined free of response-bias) would enhance the correlation between face memory and the other WMS-iii visual memory subtests. The fourth study compared performance of patients with Alzheimer's disease, Huntington's disease, Korsakoff's syndrome and demographically matched controls on the new face memory scores. Results of these experiments did not show higher correlation values with Faces I and other WMS-iii visual memory measures. One possibility is that basic cross-racial (CR) facial discriminatory biases may contribute to the low correspondence between the Faces I and similar subtests (Holdnack & Delis, 2004).

Facial recognition processing is generally measured by the ability to accurately identify facial properties (e.g., emotional expression) and individuals themselves. Humans are particularly adept at processing facial stimuli as part of an evolutionary history that relied upon the ability to discriminate reciprocating and threatening social affiliates (O’Bryan & McCaffrey, 2006). The ability to recognize individuals varies depending on the racial category of the target stimuli and the respondent being tested (O’Bryant& McCaffrey, 2006). This is sometimes referred to as cross-racial identification bias or CRIB (Jenkins, Lavie, & Driver, 2005; O’Toole, Deffenbacher, Valentin, & Abdi, 1994; Rhodes, Ewing, Hayward, Maurer, Mondloch, & Tanaka, 2009). Mitigating factors, such as race, can be seen to have instrumental effects on overall recognition patterns by influencing a person’s memory encoding processes (Massaro, & Ellison, 1996; Marcon, Susa & Meissner, 2009; Slone, Brigham, & Meissner, 2000). This effect was observed in a study conducted by Walker and Tanaka (2003), where the CRIB factor was discovered to be present during early stages of perceptual encoding, e.g., during the formation of working short-term memory. In this study, own-race advantage was found when Asian participants more accurately detected differences in other Asian faces rather than Caucasian faces. Comparatively, Caucasian subjects showed this same advantage for other Caucasian faces.

Facial recognition scales, a form of visual recognition test, have been utilized for decades with multiple psychological test batteries in order to assess memory function (O'Bryant, Hilsabeck, McCaffrey & Gouvier, 2001). One of the most familiar findings associated with facial recognition is the cross-race recognition deficit whereby subjects have difficulty recognizing the faces of members of a race different than their own (Levin, 2000). Likewise, racial identification represents an extensive area of research in the field of psychology and is generally defined as the ability for an individual to visually identify the ethnic race of another individual. Although studies involving facial recognition and racial identification are utilized extensively, the above noted results highlight the problematic limitations of measuring racial identification. Simply speaking, cross-racial identification is poorer than intra-racial identification.

There is considerable research that shows individuals are more likely to recognize faces of their own race than those of other races (Cross et al., 1971; Horry & Wright, 2008; Malpass & Kravitz, 1969; Marcon et al., 2009; Rhodes et al., 2009) and this bias appears to occur equally in both Caucasian and African American subjects (Bothwell, 1989). The possibility of racial biases in neuropsychological test materials has received increasing attention in the past 30 years, largely due to the introduction of visual recognition tests, mainly those involving photographs. For example, in a study conducted by Cross, Cross and Daly (1971), three hundred equally distributed African American and Caucasian subjects were evaluated for the recognition of 12 photographed faces. The photographed faces consisted of African American and Caucasian males and females of various ages. Caucasian subjects were better able to recognize other Caucasian faces more frequently than the African American faces. Further analysis of subject background led the researchers to conclude that cross-race (CR) recognition was moderated by racially integrated interactions prior to study participation. This moderated variable of other-race exposure is referred to as the Contact Hypothesis, and is noted to account for differences in CR identification and recognition processing. With no method for standardizing the amount of contact a subject has with other races prior to taking a neuropsychological assessment, this factor will always be present (Rhodes et al., 2009). Interpretations of the collected data were examined in light of the differences between participants' cross-racial interactions during the time in which testing occurred in order to control for mitigating variables (summer and autumn of 1969) (Cross et al., 1971). During this time period, American blacks, through work and television, were all but assured exposure to whites; however, the situation for the majority of white subjects was reversed. Differences in exposure rates theoretically accounted for the results they obtained (Cross et al., 1971). The Cross et al. (1971) study was one of the first to systematically explore CR face recognition. Their results are consistent with the pioneering research for CR recognition conducted by Malpass and Kravitz (1969), in which Caucasian and African American students showed advanced recognition ability for faces of their own race in comparison with faces of the other race. Later studies, like the one performed by Horry and Wright (2008), continue to contribute insight into the underlying foundations of CRIB.

In a 2008 study, Horry and Wright linked contextual memory to facial recognition deficits. It is purported that people are able to recognize and discriminate faces of the different and same race more readily if contextual information is supplied (Horry & Wright, 2008; Levin, 2000; Ng & Lindsay, 1994). These results were in accord with data found in previous studies (i.e., Hintz & Pezdek, 2001; Walker & Tanaka, 2003) showing perceptual processing of other-race faces to be mitigated by cross-racial exposure. These studies, as well as others (i.e., Bothwell et al., 1989; Rhodes et al., 2009), suggest that heightened discriminatory accuracy (i.e., the ability to differentiate between detailed facial features) are a result of cross-racial interaction and exposure.

Cross-racial interaction and exposure theories notwithstanding, researchers have also attempted to discover to what extent is there a biological basis for cross-racial face recognition deficits. Several studies have found cross-racial face recognition in early stages of face recognition encoding (Lindsay, Jack, & Christian, 1991; Papesh & Goldinger, 2009; Walker & Hewstone, 2006). In 1991, researchers sought to test the race-specific perceptual expertise hypothesis which states that the CRIB effect reflects race-related differences in perceptual expertise (Lindsay, et al., 1991). According to this theory, featural and configural properties of faces that support recognition differ depending on race, such that people develop specialized expertise at processing faces of particular races (Lindsay, et al., 1991). Most often, same-race recognition expertise is most adeptly developed (see Brigham & Malpass, 1985). This 1991 study was designed to provide a more direct test of the perceptual expertise hypothesis by testing an equal sample of male and female African American and White undergraduate students. Researchers were able to link race of sampled-subject with differences in same-race vs. other-race recognition, in accord with similar studies (e.g., Goldstein & Chance, 1985; Rhodes, Brake, Taylor, & Tan, 1989). In addition to finding an other-race effect, it was demonstrated that perceptual skills also play a role. If perceptual skills of an individual mitigate the recall ability with regards to cross-racial identification, then researchers may have to evaluate recall/recognition deficits on a case-by-case basis. It was highly recommended that further research into the biological underpinnings of cross-racial identification be undertaken, and that an objective measure of facial similarity be used when interpreting results.

Another study on point with the theory of underlying biological factors in cross-racial identification was conducted by Papesh and Goldinger (2009). Papesh and Goldinger, using previously published findings (e.g., Lindsay, et al., 1991; Walker & Hewstone, 2006), evaluated 300 participants over 6 experiments in an attempt to test perceptual processing intervals respective to racial category of faces. It was theorized that, even though CRIB effect is typically observed in tasks which require long-term memory, research suggests the effect can be seen early in face encoding, that is effects can be seen during the short term phase of processing a face to memory. Results of this study were somewhat surprising, showing an emergence of the other-race effect in retention and retrieval deficits, rather than in differences in immediate perceptual processing. Implications of this research will be seen in the not-to-distant future, with more and more researchers looking for biological foundations of memory deficits.

Although there has been an abundance of published literature on the assessment of deficits in CR facial recognition, there has been less focus on the assessments themselves. It is not uncommon for those in the scientific communities and, even more so, for those not trained in experimental methodology, to take for granted the assessment measures by which research is evaluated. Several research studies have shown once thought-to-be reliable assessment tools to be flawed, leading to potential reliability and validity issues. Many of the inventories used for facial discrimination and facial memory research are subject to the aforementioned problems. In addition to the WMS-iii, other tests employed to examine recognition memory include the Warrington Recognition Memory Test (RMT; Warrington, 1984), the Memory Assessment Scales (MAS; Williams, 1991), and the Rivermead Behavioral Memory Test (RBMT; Wilson, Cockburn, & Baddeley, 1985). Two potential problems in the previously mentioned tests, save for the WMS-iii, are that they utilize Caucasian photographs only and/or rely on black and white pictures (O'Bryant & McCaffrey, 2006). If the inventory testing material is not equally distributed by race, as were the cases in the aforementioned batteries, then all interpretations of results are subject to be invalid. For example, a testing inventory that includes a disproportionate amount of White/Caucasian stimuli will provide White/Caucasian subjects with a recognition advantage over other race participants (O'Bryant & McCaffrey, 2006). Additionally, for the scales that do not employ strictly Caucasian-raced sample pictures, there is the issue of racial ambiguity as a result of non-colored stimuli. If the sample pictures used for testing purposes are presented only in black and white, racial differentiation of stimuli becomes much more difficult and may lead to inaccurate test results (O'Bryant & McCaffrey, 2006).

The current study was designed to determine whether an unequal distribution of racial stimuli exists within the WMS-iii Faces I subtest. The Faces I subtest employs the use of pictured face stimuli in order to determine some aspects of memory function. Despite its widespread application (e.g., Holdnack & Delis, 2004; Migoya et al., 2002; Opasanon, 2008) studies using the subtest did not consider the role that target stimuli and responders' race plays in facial recognition. It should be noted that certain studies have empirically shown a correlation between race and a bias in facial recognition patterns (Bothwell, 1989; Cross et al., 1971; Ng & Lindsay, 1994). Based on the consistent findings of prior research studies (e.g., Cross et al., 1971; O'Bryant et al., 2001; Rhodes et al., 2009) and a pre-evaluation of the WMS-iii, it is believed there exists potential biases in the Faces I subtest, with regards to an unequal distribution of racially representative faces in the testing materials. In order to examine this hypothesis, the current research utilized the operational definitions and descriptions used in the introductory paragraphs with regards to facial recognition and race.

Method

Participants

Eighty-eight University of North Florida students (61 White/Caucasian, 11 Black/African American, 6 Hispanic, and 10 Other) participated in two separate baseline studies. Participants received no benefits or compensation for completion of the baseline study, aside from extra-credit obtained from instructors of undergraduate and graduate courses. Neither gender nor age of participants was recorded, as it was not needed for analysis purposes, however, modal age was approximately 21 years of age.

Materials and Procedure

For the current research, the Faces I subtest was utilized in a baseline capacity for perceived ethnicity amongst the target and interference faces contained within the subtest. The Faces I subtest conforms to a recognition paradigm to measure immediate and delayed memory. The subtest administration begins with the examiner exposing a series of 24 target faces, one at a time, for 2 seconds each and asking the participant to remember each face. A second series of 48 faces (24 interference faces and the original 24 target faces) is then shown to the participant. The participant is told to identify each face as either one they were asked to remember (target face) or as a new (interference) face (Wechsler, 1997). The procedures in conducting the Faces 1 subtest are essentially those involved in carrying out typical recognition memory experiments in the laboratory. It seems to be, however, that the basis for the Faces 1 subtest may not necessarily be grounded in the literature on human memory. This possibility was a key consideration in designing the current study.

Baseline Testing.

With this in mind, for the current research, the Faces I subtest was used strictly in establishing a baseline for perceived ethnicity amongst the target and interference faces contained within the subtest. In a balanced design, participants were randomly divided into one of the two baseline groups, A (target faces) or B (interference faces) and administered the subtest. All participants were given paper surveys with blank numbered slots, 1 through 24, and asked to classify each face into one of four ethnic groups: White/Caucasian, Black/African American, Hispanic or Other. No identifying marks were made on participant surveys, so as to maintain confidentiality. Participants were given unlimited time to choose the ethnicity of each face shown, and upon completion of the survey, were asked to indicate, on the top of the paper, in which ethnic group they belonged.

Letters of information and consent were distributed to participants prior to data collection sessions. Only participants who gave consent and completed the survey were included in final analysis. The primary experimenter administered the Faces I subtest to participants (no more than 2 at a time) in a single session. Collection days were pre-divided into either a target face collection session or an interference face collection session.

Collection type was established prior to participant sign-up and participant group assignments were strictly random. Target group days and interference group days alternated in order to counterbalance collections. Thus, no target group sessions were collected consecutively. If a target face session occurred on any given day, then the following collection day would be an interference face session.

For operational purposes, this study used a Kappa statistic category chart to evaluate strength of agreement between participants. Only agreement levels of substantial (.61-.80) or nearly perfect (.81-1.0) were deemed sufficient in order to determine agreement of race (Table 1; Landis & Koch, 1977). All faces with rater agreement levels of .60 and below were considered race-ambiguous (RA).

Results

Preliminary analyses of frequency distributions between race of participant and chosen race of pictured stimuli found that race of participant did play a role in the racial identification of pictured stimuli, but only in select cases. These select pictured stimuli are considered to be race-ambiguous (RA). Further analysis of the select RA cases revealed inconsistencies amongst the within group agreement rates. Distributions examining group agreement were also analyzed in order to determine the rate of racial identification agreement amongst participants of the same race. These inconsistencies and agreements will be discussed later with regard to race-dependent stimuli and its implications.

To test the hypothesis that an unequal distribution of racially representative faces exists in the WMS-iii Faces I subtest, frequency distributions were examined among percentage agreements of subjects and the racial identification of the 48 stimuli faces (24 target faces/group A and 24 interference faces/group B). Utilizing the Kappa statistical matrix, Group A (target faces) yielded an agreed racial distribution of: 9 (38%) White/Caucasian, 4 (17%) Black/African American, 5 (21%) Hispanic, 3 (13%) Other and 3(13%) RA. Group B (interference faces) yielded an agreed racial distribution of: 13 (54%) White/Caucasian, 1 (4%) Black/African American, 3 (13%) Hispanic, 1 (4%) Other and 6 (25%) RA. When combining both Group A and Group B, a percentage agreement rate of racial identification shows a 48% agreement of White/Caucasian faces amongst the 48 stimuli faces. Remaining racial identification agreements resulted in 10% Black/African American, 17% Hispanic, 8% Other and 19% RA.

A chi-squared test was used to compare the above results with the expected population for the United States of America (Table 2). It was theorized that perhaps the publishers of the Faces I subtest were not equally distributing the racial categories in their test because they were attempting to meet expected population rates. However, results showed this to not necessarily be accurate as expected number of 24.96 (64%) White/Caucasian, 5.07 (13%) Black/African American, 3.51 (9%) Hispanic, and 5.46 (14%) Other (http://www.census.gov/, retrieved June 19, 2011) with a statistical outcome X^2 (3, N = 88) = 6.486, p =.09. Racially ambiguous values were not included in the chi-squared calculations due to the inability to get a RA expected percentage of the population.

General Discussion

The increasing melting pot that typifies the population of the United States makes the need for neuropsychological testing instruments to be created for use with diverse populations a high priority. The Wechsler Memory Scale-iii is the latest version of a standardized memory inventory that is employed as the medium by which to accurately and reliably detect memory deficit in the global population. However, this scale was created on the foundation of an inventory that is long outdated in terms of cultural sensitivity. The original Wechsler Memory Scale was developed on the pretext that there were no existing biases in how individuals access memory. Furthermore, at the time of inception, the WMS did not utilize picture stimuli (Wechsler, 1997). The WMS-iii is the first Wechsler testing inventory to incorporate pictorial stimuli, however, the developers did not account for the CRIB effect when including the test pictures. This flaw in the test construction leaves the WMS-iii vulnerable to validity issues as well as reliability criticism.

The results of this baseline study demonstrated an unequal racially representative distribution among the experimental materials. Using a simple frequency distribution, a greater allotment of perceived White/Caucasian faces in both the 24 target faces as well as the 24 interference faces was revealed (Figure 1). Further, initial findings bring to light a race-dependent component amongst some of the Faces I subtest inventory pictures. This presents an issue at the core of the inventory that will undoubtedly affect scoring validity, and the overall assessment. Race-dependent inventory pictures will not have an agreed upon race, but rather will “change” race depending on the race, and possible gender, of the sample participant. The presence of race-dependent stimuli is a prime example of a cultural bias that contributes to inaccuracy issues with the WMS-iii.

Primary findings also yielded an unanticipated, yet important realization. The frequency distribution of perceived race faces in both Target and Interference groups resulted in unequal distributions as well. The distribution varied as much as double with regards to perceived race in the Target Group versus the same perceived race in the Interference group. Simply speaking, an unequal distribution within in each group may lead to an underlying bias in establishing the baseline study itself. Without a correction to this issue, the memory scale at its very core will yield inaccurate results. This unexpected result makes the baseline studies in the current research all the more important.

There are several important limitations of this study that are important to mention. One is the fact that there were time constraints for data collection, which subsequently led to a limited sample size. Data collection was limited to a six month window of participant recruitment, which proved to be less than what was ideally needed in order to recruit a sufficient number of participants. Future researchers would be wise to allow for at least a nine month time frame, if not longer, for subject recruitment. The limited sample size, though sufficient for the current study, is not comparable to the sample sizes utilized in prior research experiments, and therefore, for a more powerful result, it is recommended to test using a sample size exceeding 100 participants. This lack of power likely influenced the chi-square analyses, which yielded a p-value of .09, a not quite statistically significant value according to conventional criteria. On its face, it would seem this value is directly related to the sample size, and therefore is correctable with a simple increase in participants.

Additionally, these baseline studies were originally meant to serve as the precursor to a second experiment. Time constraints required a modification of the overall experiment model, and consequently established the current study as the lone, primary experiment. Had the current study been the primary experiment all along, subjects would have been evaluated using a repeated-measures design, rather than the balanced method. This could have led to an increase in the overall sample size, but definitely would have produced greater power and probably heightened external validity of the experiment.

One final noteworthy limitation is the population pool from which the sample was taken. This sample consisted of all undergraduate and graduate college students. College students find themselves in a diverse population on a daily basis and the likelihood of cross-racial (CR) interaction is very high. With that being said, there was no way to control for the level of prior CR interaction amongst sample subjects, and therefore, racial familiarity may have been a confounding factor that could have influenced the present results. As mentioned earlier, prior research refers to CR interaction as a moderating variable when evaluating CR recognition. It is debatable whether the CR interaction in the current study, or previous research, is the true underlying factor for observed CRIB effects. As stated in the introduction, researchers have also found evidence of a biological underpinning that may influence processing levels and the means in which cross-racial recognition is determined. If a biological factor is indeed determined, then the current limitation of the sample pool would not be particularly problematic.

Recognition of faces is a complex perceptual achievement with practical applications ranging across the spectrum (Cross et al., 1971). The results of the current study, similar to previous research, have implications for multiple fields of research, most apparent of which are in the criminal justice system (Bothwell, et al., 1989; Hintz & Pezdek, 2001; Horry & Wright, 2008; Smith, et al., 2004; Walker & Tanaka, 2003). Since the WMS-iii is a clinical measure used, for instance, in prison systems by psychologists in order to determine anything from competency to whether a defendant is legally sane, it is vital that it be accurate and reliable.

Facial recognition is a hot topic in the criminal justice system due to its direct relationship with eyewitness testimony. Researchers conducting studies on reliability and accuracy of eyewitness lineup identification and its influence regarding subsequent testimony have repeatedly warned the justice system of problems with eyewitness identification evidence (Wells & Olson, 2003). Cutler and Penrod, in a 1995 publication, noted that there had been over 2,000 scientific investigations on the reliability of eyewitness identification and the majority of these studies had yielded the same basic results (e.g., Cutler & Penrod, 1995). Since that publication, at least another 1,000 studies have been conducted, showing the same basic pattern of results (Slobogin, Rai, & Reisner, 2009). Eyewitness testimony, regardless of age or race of the witness, is often unreliable and inaccurate. This point has received even more attention since the early 90's, when the organization known as The Innocence Project was established. The sole purpose of this organization is to exonerate wrongfully convicted persons through criminal justice reforms and DNA testing. It is this organization that is responsible for one of the most publicly noted sources of eyewitness misidentification research. See (http://www.innocenceproject.org/docs/Eyewitness_ID_Report.pdf.).

In their "Reevaluating Lineups" report, The Innocence Project reported that as of 2010, of the 230 people that have been exonerated through DNA testing, 179 (75%) of those were convicted on primarily eyewitness testimony, and 53% of the misidentifications involved cross-racial misidentification (West, 2010). This research does not focus on why cross-racial identification deficits occur, but it does show real-world application of these deficits and how serious the consequences can be when cross racial identification research is not considered or is inaccurately used. It should be noted that this report only used a sample population of persons convicted and sent to prison for 12 years or more. Generalizing to all crimes, this report suggests how widespread and significant cross-racial identification biases are, and how significant accurate testing measures and research should be regarding the general population. Fortunately, certain states have implemented legislation in order to account for unreliability in eyewitness testimony, however, this legislature is more of a broad ruling on eyewitness testimony, and does not directly focus attention on the cross-racial biases that contribute to eyewitness testimony inaccurate (Slobogin, Rai, & Reisner, 2009).

In a study conducted by Bothwell et al. (1989), a meta-analysis of 14 studies was examined in order to determine to what extent, if any, Black and White subjects are vulnerable to an own-race bias. Consistency was found across all studies, indicating that memory for own-race faces are superior to the memory for other-race faces. Both Black and White subjects exhibited a bias for own-race faces in 79% of the samples. Following the results of this, and other similar research, the WMS-iii shows an empirical problem with regards to reliability. The chi-squared analysis revealed a racial distribution similar to what a researcher might find in expected values of racial distribution in the United States, but that is irrelevant. When administering the WMS-iii Faces I subtest, the subjects are not racially distributed according to the most recent census. The subject will be of a single race and, whether they are White/Caucasian, Black/African American, Hispanic or Other; that person should have an equal opportunity for recognition.

O'Bryant and McCaffrey (2006) conducted the only study to date, that I am aware of, that mentions a potential unequal distribution of racially representative photos in the WMS-iii Faces I subtest. It is noted in their research that a potential bias in the WMS-iii is known to the authors of the scale, however, to date there has not been a published study that explicitly examines this subtest. Furthermore, it is implied that research directly showing a link between the WMS-iii and an unequal distribution will motivate the publishers of the scale to correct the issue (O'Bryant & McCaffrey, 2006). It was an aim of the current study to supply some of that motivation, if not to the publishers of the WMS-iii, then at least to other researchers to further investigate.

The WMS-iii is not only culturally biased based on the current findings, generalized to the U.S. population, but is invalid in any country where White/Caucasian is not the majority. Aside from Europe and North America, this makes the WMS-iii significantly questionable with regards to validity of the measure’s results. Perhaps further investigations can yield a means to develop an international testing battery that can benefit all populations, regardless of racial distribution. With the United States and the world being such a melting pot of races, it is a wonder this has not been developed to date.

As mentioned in the introductory literature review, the WMS-iii has been the subject of little research, and the effects of race on the perceived memory scores of the Faces I subtest has received even less analysis. Results from the current research show an unequal distribution of race in the subtest stimulus inventory. This, accompanied with subject test data, suggests a bias in the Wechsler Memory Scale-iii Faces I subtest for subjects conducting cross-racial recall or recognition. Without a standardization of the picture stimuli to account for the race-dependent factor, as well as other variables such as racial majority in population sample, this scale will continue to yield invalid and possibly detrimental results. As previously stated, applicability of research in the area of neuropsychological evaluation and validity of testing measures has an immediate, real world impact in various fields of study. Severe consequences, such as the loss of liberty and life (see applicability of cross-racial misidentification in eye-witness testimony) have been directly related to this area of research, making the need for reliable and valid testing measures all the more prevalent. Without accurate assessment tools, it is impossible to find the underlying methodologies that can adversely affect contemporary society. It is the hope that this, as well as future research, will influence the publishers of the WMS-iii to correct any flaws and underlying bias contained therein.

In an updated note, since the beginning of the current experiment, the WMS-iv has been developed and distributed, and the Faces I subtest has been eliminated from the primary testing material. Reasoning provided by publishers for discarding the Faces I subtest include: “sensitivity primarily to disorders associated with social perception impairment (e.g., schizophrenia, autism, Asperger’s syndrome); issues with relatively low reliability (due to a high guess rate); and clinical sensitivity issues with floor problems (random responding resulted in a low average score) (WMS-iii to WMS-iv:Rationale for Change, 2012).” It is unfortunate that the publishers did not address the underlying problem of validity in the Faces I subtest. I suppose eliminating the subtest from the primary testing material solves the issue of unequal distribution, however, it does not solve the lack of standardization or oversight in test construction that was found in the Faces I. This leads a researcher to wonder if the Faces I is not alone, with regards to subtests with major underlying validity issues in the WMS. Future research may want to focus on the new “updated” subtests in the WMS-iv before taking on face value the validity of the results. If a simple issue concerning racial distribution can go uncorrected by the publishers, what more complex underlying issues have slipped by, thus invalidating test results?

Appendix A (Measures)

Choose Which Number Fits the Ethnicity of the Corresponding Picture Best

1	2	3	4
White/Caucasian	Black/African American	Hispanic	Other

1.______________

2.______________

3.______________

4.______________

5.______________

6.______________

7. ______________

8. ______________

9. ______________

10.______________

11. ______________

12. ______________

13. ______________

14. ______________

15. ______________

16. ______________

17. ______________

18. ______________

19.______________

20.______________

21.______________

22.______________

23.______________

24.______________

References

Anastasi, J. S. & Rhodes, M. G.. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review, 12*(6), 1043-1047.

Bothwell, R. K., Brigham, J. C., & Malpass, R. S.. (1989). Cross-racial identification. *Personality and Social Psychology Bulletin, 15*(1), 19.

Chiroro, P., & Valentine, T.. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *Quarterly Journal of Experimental Psychology, 48A*, 879-894.

Cross, J. F., Cross, J., & Daly, J.. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception & Psychophysics 10*(6), 393-396.

Fiset, D., Blais, C., Gosselin, F., Bub, D., & Tanaka, J.. (2008). Potent features for the categorization of Caucasian, African American, and Asian faces in Caucasian observers. *Journal of Vision, 8*(6), 258.

Franzen, M. D.. (2000). Reliability and validity in neuropsychological assessment (second edition). New York, N.Y.: Plenum Publishers.

Goldstein, A. G., & Chance, J. E.. (1985). Effects of training on Japanese face recognition: Reduction of the other-race effect. *Bulletin of the Psychonomic Society*, 23, 211-214.

Hilsabeck, R. C., Dunn, J. T., & Lees-Haley, P. R.. (1996). An empirical comparison of the Wechsler Memory Scale-Revised and the Memory Assessment Scales in measuring four memory constructs. *Assessment*, *3*(4), 471-422.

Hinz, T., & Pezdek, K.. (2001). The effect of exposure to multiple lineups on face identification accuracy. *Law and Human Behavior, 25*(2), 185.

Holdnack, J. A. & Delis, D. C.. (2004). Parsing the recognition memory components of the WMS-III face memory subtest: Normative data and clinical findings in dementia groups. *The Journal of Clinical and Experimental Neuropsychology,* 26(4), 459-483.

Horry, R., Wright, D., & Tredoux, C.. (2010). Recognition and context memory for faces from own and other ethnic groups: A remember-know investigation. *Memory & Cognition, 38*(2), 134-41.

Jenkins, R., Lavie, N., & Driver, J.. (2005). Recognition memory for distractor faces depends on attentional load at exposure. *Psychonomic Bulletin & Review, 12*(2), 314-320.

Landis, R. J. & Koch, G. G.. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Levin, D. T.. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General, 129*(4), 559-574.

Lindsay, D. S., Jack, P. C., Jr., & Christian, M. A.. (1991). Other-race face perception. *Journal of Applied Psychology*, 76, 587–589.

Malpass, R. S., & Kravitz, J.. (1969). Recognition for faces of own and other race. *Journal of Personality & Social Psychology, 13,* 330-334.

Marcon, J., Susa, K., & Meissner, C.. (2009). Assessing the influence of recollection and familiarity in memory for own- versus other-race faces. *Psychonomic Bulletin & Review, 16*(1), 99-103.

Massaro, D. W., & Ellison, J. W.. (1996). Perceptual recognition of facial affect: Cross-cultural comparisons. *Memory & Cognition, 24*(6), 812-822.

Metzger, M. M.. (2006). Face distinctiveness and delayed testing: Differential effects on performance and confidence. *The Journal of General Psychology, 133*(2), 209-16.

Ng, W., & Lindsay, R. C. L.. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross - Cultural Psychology, 25*(2), 217.

Noor, A., & Najam, N.. (2009). Comparative study of memory deficits in younger and older adults. *Pakistan Journal of Psychological Research, 24*(3/4), 145-158.

O'Bryant, S. E., Hilsabeck, R. C., McCaffrey, R. J., & Gouvier, W. D.. (2001). The recognition memory test examination of ethnic differences and norm validity. *Archives of Clinical Neuropsychology, 18*, 135-143.

Opasanon, N.. (2008). A study of the Wechsler Memory Scale-Third Edition Abbreviated in the elderly dementia patients. Master's thesis, Mahidol University, Thailand.

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H.. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition, 22*(2), 208-224.

Papesh, M., & Goldinger, S.. (2009). Deficits in other-race face recognition: No evidence for encoding-based effects. *Canadian Journal of Experimental Psychology, 63*(4), 253-262.

Parra, M.A., Abrahams, S., Fabi, K., Logie, R., Luzzi, S., & Della Sala, S. (2009). Short-term memory binding deficits in Alzheimer's Disease. *Brain, 132*, 1057-1066.

Parra, M.A., Abrahams, S., Logie, R., Mendez, L.G., Lopera, F., & Della Sala, S.. (2010). Visual short-term memory binding deficits in Familial Alzheimer's Disease. *Brain, 133*, 2702-2713.

Pauker, K.. Not so black and white: The impact of motivation on memory for racially ambiguous faces. PhD. dissertation, Tufts University, United States -- Massachusetts. Retrieved April 13, 2010, from Dissertations & Theses: A&I.(Publication No. AAT 3354706).

Rhodes, G., Brake, S., Taylor, K., &Tan, S.. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology*, 80, 313-331.

Rhodes, G., Ewing, L., Hayward, W., Maurer, D., Mondloch, C., & Tanaka, J.. (2009). Contact and other-race effects in configural and component processing of faces. *British Journal of Psychology, 100*(4), 717.

Searcy, J. H., Bartlett, J. C., & Memon, A.. (1999). Age differences in accuracy and choosing in eyewitness identification and face recognition. *Memory & Cognition, 27*(3), 538-552.

Slobogin, C., Rai, A., & Reisner, R.. (2009). *Law and the Mental Health System, 5d,* 555.

Slone, A. E., Brigham, J. C., & Meissner, C. A.. (2000) Social and cognitive factors affecting the own-race bias in whites. *Memory & Cognition, 28,* 71-84.

Smith, S. M., Stinson, V., & Prosser, M. A.. (2004). Do they all look alike? An exploration of decision-making strategies in cross-race facial identifications. *Canadian Journal of Behavioral Science*, *36*(2), 146-154.

Sporer, S. L.. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law, 7*, 36-97.

Tanaka, J., & Pierce, L.. (2009). The neural plasticity of other-race face recognition. *Cognitive, Affective and Behavioral Neuroscience, 9*(1), 122-131.

Ullman, M. T.. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*, 231–270.

Walker, P. M., & Hewstone, M.. (2006). A perceptual discrimination investigation of the own-race effect and intergroup experience. *Applied Cognitive Psychology*, 20, 461–475.

Walker, P. M., & Tanaka, J.. (2003). An encoding advantage for own-race versus other-race faces. *Perception, 32*, 1117-1125.

Warrington, E. K.. (1984). *Manual for Recognition Memory Test.* Windsor, England: NFER-Nelson.

Wechsler, D.. (1997). *The Wechsler* Memory Scale-III *manual*. San Antonio, TX: The Psychological Corporation.

Wechsler, D.. (2004). *The Wechsler* Memory Scale-III *manual*. San Antonio, TX: The Psychological Corporation.

Wechsler, D.. (1997). Wechsler *Adult Intelligence Scale-3rd ed: Manual.* San Antonio, TX: Psychological Corporation.

Wells, G. L. & Olson, E. A.. (2003). Eyewitness testimony. *Annual Review of Psychology, 54*, 277-295.

West, E. M.. (2010), Reevaluating lineups: Why witnesses make mistakes and how to reduce the chance of a misidentification. Retreived October 10, 2011from, http://www.innocenceproject.org/docs/Eyewitness_ID_Report.pdf.

Wickham, L. H. V., Morris, P. E., & Fritz, C. O.. (2000). Facial distinctiveness: Its measurement, distribution and influence on immediate and delayed recognition. *British Journal of Psychology, 91*(1), 99-123.

Wilson, B., Cockburn, J., & Baddeley, A.. (1985). *The Rivermead Behavioral Memory Test: Manual*. Bury St. Edmunds, England: Thames Valley Test Co..

Wixted, J. T.. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114(1)*,152-176.

WMS-iii to WMS-iv:Rationale for Change. Retreived February 21, 2012 from http://psychcorp.pearsonassessments.com/hai/images/Products/WMS-IV/WMS-RationaleforChange.pdf.

Yonelinas, A.P., Dobbins, I., Szymanski, M.D., Dhaliwal, H.S., & King, L.. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition*, *5*, 418-441.

Table 1.

Value of Strength of Inter-rater Agreement:
(Landis & Koch, 1977).

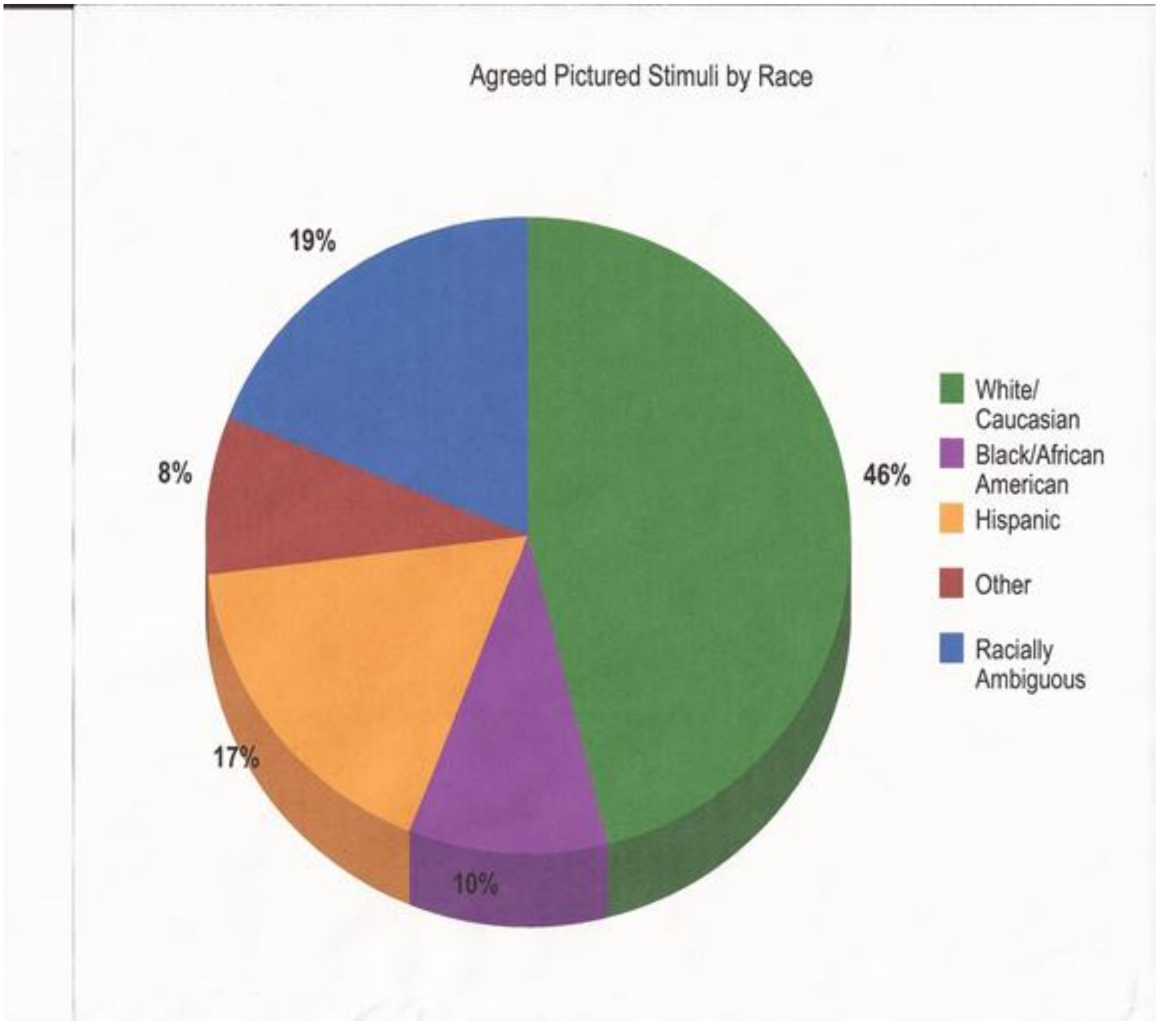
< 0.20 Poor
0.21 - 0.40 Fair
0.41 - 0.60 Moderate
0.61 - 0.80 Good (Substantial)
0.81 - 1.00 Very good (Nearly Perfect)

Table 2.

Observed vs. Expected Racial Distribution

Row #	Category	Observed	Expected	%Expected
1	White/Cauc.	22	24.96	64.000%
2	Black/A.Am.	5	5.07	13.000%
3	Hispanic	8	3.51	9.000%
4	Other	4	5.46	14.000%

Figure 1.


Agreed Pictured Stimuli by Race
19%
8%
46%
17%
10%
White/ Caucasian
Black/African American
Hispanic
Other
Racially Ambiguous

Vita

ADAM LESS

EDUCATION

Drake University Law School — Des Moines, IA
Juris Doctor Candidate — May 2013
Affiliations: Drake Law Negotiations Team (2011-2012)
Delta Theta Phi Law Fraternity (2010 –2012)
Research: Westlaw/LexisNexis Certified (2010-2012)

University of North Florida — Jacksonville, FL
Master of Arts, General Psychology — February 2011
Bachelor of Science, Psychology — May 2008
Honors: Psi Chi National Honor Society, Co-President (2009)
Research Honors: Recipient, Undergraduate Research Grant for studying relations between preschool-aged children and cognitive development – Primary Investigator, data collection, analysis, write-up, article research, summarization, and information presentation
Leadership: Criminal Justice Association (2005 – 2010); Boxing Club, Vice President (2005 – 2006); Pi Kappa Phi National Fraternity (2005 – 2010), Risk Management Chair, Pledge Class Vice-President, P.U.S.H. Chair
Affiliations: American Psychology Association Student Affiliate (2006 – 2009)

EDUCATIONAL EXPERIENCE

University of North Florida — Jacksonville, FL
Supervised Research — 2008-2011
*Cultural Biases in the WMS-iii

University of North Florida — Jacksonville, FL
*Learning and Cognition — Spring 2010
*Advanced Human Development — Spring 2010
*Advanced Social Psychology — Fall 2009
*Colloquium in Psychological Research — Fall 2009
*Evolutionary Psychology — Fall 2009
*Research, Design and Analysis — Spring 2009
*Advanced Biopsychology — Spring 2009
*Advanced Personality Theories — Fall 2008
*Statistical Methods in Social Sciences — Fall 2008
*Seminar in Psychological Sciences — Fall 2008

*Colloquium in Psychological Sciences	Fall 2008
*Experimental Social Psychology	Spring 2008
*Planning and Control	Spring 2008
*Psychological Testing	Spring 2008
*Cognitive Psychology	Spring 2008
*Psychobiology	Fall 2007
*Personal Health Issues and Problems	Fall 2007
*Experimental Cognitive Psychology	Fall 2007
*History of Psychology	Spring 2007
*Research Methods/Lab in Psychology	Spring 2007
*Introduction to Counseling	Fall 2006
*Major in Psychological Seminar	Fall 2006
*Theories of Personality	Summer 2006
*Stress Management	Summer 2006
*Psychology of Abnormal Behavior	Summer 2006
*Personal Growth	Summer 2006
*Foundations of Child Psychology	Spring 2006
*Social Psychology	Fall 2006
*Lifespan Development Psychology	Summer 2005
*Introduction to Psychology	Fall 2002

PUBLICATIONS AND PRESENTATIONS

*Perez, S., & Less, A. Relations between Preschool Children's Observed Planning, Inhibitory Control, Teacher-Rated Temperament, and Letter Recognition. Osprey Journal of Ideas and Inquiry. (February 2009).
*Poster session presented at: the biennial meeting of the Society for Research in Child Development. Denver, CO (April 2009) and the Annual Student Research Symposium. Jacksonville, FL (March 2008).

ADDITIONAL INFORMATION

Skills: Computer Efficient - typing (75 wpm), APA Formatting, Powerpoint, MS Word, Excel, SPSS, ParScore

Areas of Emphasis: Forensic Psychology, Neuropsychology and Assessment