AN INVESTIGATION OF MALINGERING AND DEFENSIVENESS USING THE SPANISH PAI AMONG SPANISH-SPEAKING HISPANIC AMERICAN OUTPATIENTS

Amor Alicia Correa, M.S.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

August 2013

APPROVED:

Richard Rogers, Major Professor
Randall Cox, Committee Member
John Ruiz, Committee Member
Vicki Campbell, Chair of the Department of Psychology
Mark Wardell, Dean of the Toulouse Graduate School
For response styles, malingering describes the deliberate production of feigned symptoms by persons seeking external gain such as financial compensation, exemption from duty, or leniency from the criminal justice system. In contradistinction, defensiveness occurs when patients attempt to downplay their symptoms of psychological impairment. Both of the aforementioned response styles can markedly affect the accuracy of diagnosis, especially on self-reports, such as multiscale inventories. As an important oversight, no studies have been conducted to examine the effect of culturally specific response styles on profile validity and the classification of malingering among Hispanic American clinical populations. The current study investigated whether the Spanish Personality Assessment Inventory (PAI) effectively distinguished between Spanish-speaking outpatient groups randomly assigned to honest, feigning, and defensive experimental conditions. In examining the results, PAI malingering indicators utilizing Rare Symptoms strategies (NIM and MAL) demonstrated moderate to large effect sizes. For defensiveness, Spanish PAI indicators also demonstrated moderate to very large effect sizes ($M_d = 1.27$; range from 0.94 to 1.68). Regarding psychometric properties, Spanish PAI validity scales, provide adequate to good data on reliability and discriminant validity. Clinical utility of the Spanish PAI increases as different cut scores are employed.
Copyright 2013

by

Amor Alicia Correa
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>vi</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER 1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Assessment Needs of Hispanic Americans and Spanish-Speaking Assessment Clients</td>
<td>1</td>
</tr>
<tr>
<td>Acculturation</td>
<td>3</td>
</tr>
<tr>
<td>Culturally-Specific Response Patterns and Other Factors Affecting Assessment with Hispanic Americans</td>
<td>6</td>
</tr>
<tr>
<td>Cultural Responses and Other Considerations for Intelligence Testing with Hispanic American Clients</td>
<td>7</td>
</tr>
<tr>
<td>Culturally Specific Response Patterns Which Affect Validity Scale Scores for Hispanic Americans</td>
<td>12</td>
</tr>
<tr>
<td>Validity of Assessment Measures for Ethnic Minority Individuals</td>
<td>13</td>
</tr>
<tr>
<td>Response Styles</td>
<td>15</td>
</tr>
<tr>
<td>Malingering</td>
<td>18</td>
</tr>
<tr>
<td>Defensiveness</td>
<td>24</td>
</tr>
<tr>
<td>The Development of Detection Strategies for Malingering and Defensiveness</td>
<td>25</td>
</tr>
<tr>
<td>Assessment of Malingering and Defensiveness</td>
<td>28</td>
</tr>
<tr>
<td>Assessment of Response Styles Using Multiscale Inventories</td>
<td>29</td>
</tr>
<tr>
<td>The MMPI-2</td>
<td>29</td>
</tr>
<tr>
<td>The PAI</td>
<td>33</td>
</tr>
<tr>
<td>The Bipolarity Hypothesis</td>
<td>38</td>
</tr>
<tr>
<td>Malingering and Defensiveness among Mexican Americans</td>
<td>39</td>
</tr>
<tr>
<td>Multiscale Inventories</td>
<td>39</td>
</tr>
<tr>
<td>MMPI-2</td>
<td>40</td>
</tr>
<tr>
<td>PAI</td>
<td>42</td>
</tr>
<tr>
<td>Spanish SIRS-2</td>
<td>45</td>
</tr>
<tr>
<td>Linguistic and Cultural Considerations when Using the Spanish PAI</td>
<td>46</td>
</tr>
<tr>
<td>Purpose of the Current Study</td>
<td>48</td>
</tr>
<tr>
<td>Research Questions and Hypotheses</td>
<td>48</td>
</tr>
<tr>
<td>Supplementary Question</td>
<td>50</td>
</tr>
</tbody>
</table>
CHAPTER 2. METHODS .................................................................................................................. 51
  Study Design ............................................................................................................................... 51
  Participants .................................................................................................................................. 52
  Materials ....................................................................................................................................... 53
    Spanish Personality Assessment Inventory (PAI; Morey, 1991) ........................................... 53
    The Acculturation Rating Scales for Mexican Americans—2nd edition (ARSMA-II; Cuellar, Arnold, & Maldonado, 1995) ............................................................... 54
    Reading Level Indicator (RLI; Williams, 2000) ........................................................................ 54
    Demographics Questionnaire ..................................................................................................... 54
  Procedure ..................................................................................................................................... 55
    Phase I ....................................................................................................................................... 55
    Phase II ..................................................................................................................................... 57
    Scenarios ................................................................................................................................... 57
    Manipulation Check .................................................................................................................... 58
    Procedure for the Exclusion of Invalid Profiles ........................................................................ 59

CHAPTER 3. RESULTS .................................................................................................................... 61
  Refinement of the Sample ........................................................................................................... 61
  Demographic Data ........................................................................................................................ 62
  Effectiveness of the Spanish PAI Validity Indicators ................................................................. 65
    PAI Validity Indicators ................................................................................................................. 65
    Utility of Spanish PAI Scales ....................................................................................................... 68
  Internal Consistency of the Spanish PAI Validity Scales ............................................................. 78
  Acculturation ................................................................................................................................ 79
  The Bipolarity Hypothesis ........................................................................................................... 80
  Effects of Clinical Symptoms on Validity Indicators .................................................................... 82

CHAPTER 4. DISCUSSION ................................................................................................................. 84
  Culturally-Specific Response Patterns and Hispanic Americans .................................................. 88
  Classification Accuracy for the Spanish PAI Feigning Indicators .................................................. 96
  Bipolarity Hypothesis for Feigning and Defensiveness .................................................................. 99
  Reliability of the Spanish PAI ....................................................................................................... 101
  Validity of the Spanish PAI for Feigning Indicators ....................................................................... 103
  Effects of Acculturation on the Spanish PAI ................................................................................. 106
LIST OF TABLES

1. Description of Response Styles ................................................................. 17
2. Description of Detection Strategies for Malingering ............................... 26
3. Description of Detection Strategies for Defensiveness ............................ 27
4. A Comparison of Male and Female Hispanic American Outpatients on Demographic Variables .......................................................... 64
5. A Comparison of Male and Female Honest Responding Outpatients on PAI Validity Indicators ................................................................. 65
6. Differences on the Spanish PAI Validity Indicators Between Honest and Feigned Presentations ................................................................. 66
7. Differences on the Spanish PAI Validity Indicators Between Honest and Defensive Presentations ................................................................. 67
8. Mean Values for INF Item Endorsement by Hispanic American Outpatients on the Spanish PAI for Honest, Malingering, and Defensive Conditions ........................................................................ 68
9. Utility of PAI Feigning Indicators for Differentiating between Likely Genuine and Likely Feigning Responders .............................................................. 70
10. Effectiveness of PAI Cut Scores for Feigning with the Exclusion of an Indeterminate Category ................................................................. 72
11. Errors in the Indeterminate Group for PAI Cut Scores on Malingering Indicators: False Alarms and False Misses at 50% Base Rate ........................................... 73
12. Utility of PAI Defensiveness Indicators for Differentiating between Likely Genuine and Likely Defensive Responders .............................................................. 76
13. Effectiveness of PAI Cut Scores for Defensiveness Scales with the Exclusion of an Indeterminate Category ................................................................. 77
14. Errors in the Indeterminate Group for PAI Cut Scores: False Alarms and False Misses at 50% Base Rate ................................................................. 78
15. Internal Consistencies and Standard Errors of Measurements (SEM) for the Spanish PAI Validity Scales ................................................................. 79
16. Acculturation as a Predictor for Scores on PAI Validity Indicators of Honest Responders ................................................................. 80
17. Pearson Correlation Matrix for Spanish PAI Validity Indicators among Hispanic American Outpatients in the Honest Condition .............................................................. 81
18. Differences on the Spanish PAI Validity Indicators for Patients Diagnosed with Only Mood Disorders in the Honest Condition ................................................. 82
19. Percent of Endorsement for PAI Ratings across Experimental Conditions .......... 90
20. A Comparison of Effect Sizes Between Honest and Feigning Conditions ............ 93
21. A Comparison of Effect Sizes Between Honest and Defensive Conditions in Clinical and Non-clinical Samples of Hispanic Americans on the Spanish PAI .. 95
22. A Comparison of Internal Consistency Determined by Alpha Coefficients Across English and Spanish PAI Studies ................................................................. 102
CHAPTER 1
INTRODUCTION
Assessment Needs of Hispanic Americans and Spanish-Speaking Assessment Clients

Currently, most American assessment measures in the field of psychology have been developed for clients proficient in English and validated on clinical samples primarily composed of European American individuals. However, the status quo is changing because of increased cultural diversity in the United States plus a greater awareness of how cultural differences affect test results and their interpretation. Clearly, assessment measures must consider the unique cultural needs of ethnic minority individuals.

These cultural considerations are particularly salient for Hispanic Americans, given the heterogeneity of their cultural backgrounds represented in the United States and compounded by the challenges with translating measures from English to Spanish. The Hispanic American population is currently the fastest-growing minority group in the United States. According to the most recent available census data the Hispanic American population of the United States grew by 43% between 2000 and 2010 (US Census Bureau, 2011a). Moreover, a large proportion of these individuals report Spanish as their primary language. In fact, across all ethnicities and cultural groups in the United States, nearly 62.1% of individuals who primarily speak a language other than English in their home identified their primary language as Spanish. Of these individuals, nearly one-third (27.7%) reported speaking English not well or not at all (US Census Bureau, 2011b). This growing Spanish-speaking subpopulation creates a
compelling need for assessment tools with norms that are reliable and valid for use with, not only Hispanic American populations, but specifically with Spanish speakers.

Importantly, however, cultural considerations for assessment practices extend far beyond a prudent recommendation. Ethical guidelines from the American Psychological Association (2002) require that psychologists working with culturally diverse populations recognize these characteristics as important factors affecting a person’s experiences, attitudes, and psychological presentation. The distinctions are especially pronounced when a person’s culture and primary language vary from the normative sample (Bersoff, 2004; Weiss & Rosenfeld, 2012). Standard 9.02 of the APA code of ethics (2002) specifically instructs psychologists to use assessment methods that are appropriate to an individual’s language preferences and to describe specific strengths and limitations of these measures when psychometric properties of a test have not been established for use with the population in question.

The current study investigates the potential effects of culture on validity indices of the Spanish Personality Assessment Inventory (PAI; Morey, 2007). While initial validation studies have been conducted for the translated measure (Rogers, Flores, Ustad, & Sewell, 1995), there remains a dearth of information regarding the effects of culturally-specific response patterns on the validity of test profiles. Furthermore, the ability of the Spanish PAI to effectively distinguish between patients reporting honestly and those under-reporting or over-reporting symptoms has yet to be systematically investigated (Fernandez et al., 2008). The following sections discuss cultural differences, response styles, and the effects of both on psychological assessment measures.
Acculturation

Variations among persons with different cultural or ethnic backgrounds impact the efficacy and clinical relevance of psychological assessments and subsequent interventions. Thus, differences in response patterns of distinct ethnic groups must be empirically researched so that they can be systematically addressed when interpreting standardized testing measures (Anastasi, 1988). To avoid dichotomous classification, levels of acculturation for members of ethnic minority groups must also be considered. Acculturation can be defined as the changes that occur in an individual's beliefs and behaviors, as a result of interaction with his or her own ethnic group as well as the new cultural group. Individuals with higher levels of acculturation have a greater understanding of the new culture and begin to accept and incorporate aspects of it into their daily lives (Wagner & Gartner, 1997). As a seminal model, Berry, Kin, Power, Young, and Bujaki (1989) proposed a two-dimensional model of acculturation, which provides a conceptual framework for the validation of culturally sensitive measures. In this model, individuals may experience differing needs to identify with both their own minority culture and with the majority culture. The individual can maintain one of four possible relationships with majority and minority cultures:

- **Assimilation**: sole identification with the majority culture
- **Integration**: identification with both cultures
- **Separation**: sole identification with the minority culture
- **Marginality**: no identification with either culture

Berry et al.'s (1989) represents a bidimensional model of acculturation, where it is possible for the individual to maintain varying degrees of affiliation with minority and
mainstream cultures. In contrast, unidimensional models of acculturation are also available, which contend that one relationship must always be stronger than the other (Gordon, 1964). In unidimensional models, individuals are generally conceptualized as relinquishing their ethnic culture, as they become more assimilated to mainstream American culture.

In both models, distinct levels of acculturation increase the variety of possible response patterns on psychological measures because differences also exist within cultures, not just between them. On this point, unidimensional models likely obscure the complexity of individual acculturation, by failing to recognize bicultural individuals who identify strongly with both cultures (Ryder, Alden, & Paulhus, 2000). However, both models emphasize the notion that all members of an ethnic minority cannot be simply grouped together when data are analyzed. How acculturation differentially affects responses to test items should also be determined when establishing new normative samples and cut scores on new or translated measures.

In psychological assessment, issues of acculturation must be considered for individuals whose primary identification is toward a different culture (i.e., the traditional orientation). Researchers and practitioners both recognize that standardized assessment measures administered to individuals who are culturally different from the normative sample can have quite different psychometric characteristics, which may lead to biased results as well as incorrect classification of individuals from different cultural groups (Dana, 2005; Marin & Marin, 1991). In large part, culturally biased assessment results occur in the United States, because interpretive norms, which were developed mostly on individuals of European American heritage, can only be considered valid for
the European American culture if no further testing is conducted with other cultures (Berry 1969, 1988, 1989; Dana, 2005). Omitting analysis of cultural variables in test development effectively forces minority individuals into the same interpretative categories as European Americans and creates a substantial possibility for misdiagnosis and misinterpretation (Dana, 1993; Todd, 2005).

Researchers find that English language measures adapted for Spanish speakers often fail to evaluate level of acculturation (Echemendia & Harris, 2004; Renteria, 2005; Salazar, Perez-Garcia, & Puente, 2007). Regarding this issue, Lucio, Durán, Graham, and Ben-Porath (2002) demonstrated the detrimental effects of failing to acknowledge cultural differences. In their study of the Minnesota Multiphasic Personality Inventory – Adolescent Version (MMPI-A) and Mexican adolescents, they found notably different cut scores were necessary for juveniles in the United States and Mexico. Specifically, Lucio et al. (2002) found a cut score of $F \geq 31$ correctly identified all feigners in their clinical sample of Mexican youths, whereas a previous study of Hispanic Americans (Stein, Graham, & Williams, 1995) utilized a much lower cut score of $F \geq 23$ to correctly identify 100% of feigners. Lucio et al. (2002) primarily attribute these differences in appropriate cut scores to cultural differences in response styles which influence scores obtained from Mexican adolescents in a distinctive manner than scores from American adolescents of Hispanic descent. They posited that clinical samples of Mexican adolescents are possibly more likely to exaggerate their symptoms or admit them more openly than Hispanic adolescents living in the United States, causing the notable disparity.
Such omissions in considering the effects of cultural differences are not limited to the MMPI versions, such as the MMPI-A. To date, the effects of acculturation on the PAI remain uninvestigated and existing studies are limited to samples of bilingual individuals. Neglect of monolingual populations limits generalizability, because they generally differ from the dominant culture to a greater extent than their bilingual counterparts (Correa & Rogers, 2010; Fernandez et al., 2008). To address this oversight, the current study examines the effects of acculturation on PAI scales in a monolingual, clinical sample.

Culturally-Specific Response Patterns and Other Factors Affecting Assessment with Hispanic Americans

Several culturally-specific response patterns have been identified throughout multicultural assessment literature (Correa & Rogers, 2010; Geisinger, 1994; Marin & Marin, 1991; Todd, 2005). Unique response patterns among minority groups that are substantially different from the normative sample generally indicate shortcomings in the measure’s generalizability for the cultural group in question (Hambleton, 2001). When applicable, population-specific response patterns should be taken into account when interpreting assessment results in order to avoid possible misdiagnosis resulting from the application of norms that are not appropriate for clients of diverse cultural backgrounds (Helms, 1992).

Additional considerations must be taken to ensure test validity when a psychological measure is translated into another language. The Test Translation and Adaptation Guidelines developed by the International Test Commission (ITC) in 1992 called for test developers and publishers to apply appropriate research methods and
statistical techniques to establish the validity of a test in each population for whom the adapted version is intended. Research results must be used to improve the accuracy of the translation/adaptation process and to identify problems in the adaptation that may render a measure inadequate for use with the intended populations. Additionally, test developers should strive to establish the equivalence of the different language versions of the test, to make them as parallel to the original as possible. Lastly, the validity of the translated version must be determined separately from that of the original measure. It should not be assumed that a translated version has acceptable validity simply because that of the original English language version is adequate (Allalouf, 2003; Anastasi, 1988). Until the reliability and validity of these assessment measures has been determined, mental health professionals should refrain from using them just as they would refrain from administering any other unvalidated measure (Allalouf, 2003; Hambleton, 2001).

What follows is a discussion of response patterns commonly displayed by Hispanic American clients on standardized assessment measures in various domains of psychological testing. Special attention is given to response patterns evident on assessment measures that have been translated from English to Spanish. Other considerations involving appropriate normative samples are also addressed.

Cultural Responses and Other Considerations for Intelligence Testing with Hispanic American Clients

For intelligence testing, researchers have long since pointed out that demographic variables such as age, gender, and culture affect an individual’s performance on cognitive tests (Kaufman & Lichtenberger, 2002). Using data from the
English language WAIS-WMS co-norming project, Heaton, Taylor, and Manly (2001) found Hispanic American individuals generally achieved lower scores than their European American counterparts when both groups were tested in English. Using standard norms, between 15 and 25 percent of Hispanic individuals were misclassified as “impaired” on the WAIS and WMS even when corrections were made for other factors such as age, gender, and level of education. In order to reduce an apparent bias in the interpretation of the measure, normative adjustments were suggested by Heaton et al. (2001). Predictably, when using the corrected cut scores, Hispanic American individuals have nearly the same likelihood of being misclassified as their European American counterparts.

Kaufman and Lichtenberger (2002) hypothesized that lower scores for Hispanic individuals on verbal measures reflect (a) unfair language demands placed on individuals for whom English is a second language, and (b) the cultural content of some verbal test items. Similarly, the Standards for Educational and Psychological Testing from the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (AERA, APA, NCME, 1999) specify that any oral or written test is also inherently a measure of an examinee’s verbal skills, whether it aims to measure this construct or not. Thus, reliance on verbal tests creates significant concerns for individuals whose primary language is not the original language of the test. In those cases, “test results may not reflect accurately the qualities and competencies intended to be measured” (AERA, APA, NCME, 1999, p. 91). In light of misclassification rates noted for Hispanic American individuals by Heaton
et al. (2001), mental health professionals should be cautious in interpreting results and should use alternate cut scores when appropriately validated.

In general, clinicians are severely limited in their choices of culturally appropriate assessment measures for Spanish-speaking clients. Test manuals of intelligence measures with Spanish translations of test items, such as the Kaufman Brief Intelligence Test Second Edition (K-BIT 2; Kaufman & Kaufman, 2004), warn that the test is “not intended to be administered in Spanish” (Kaufman & Kaufman, 2004, p.1) due to a lack of research testing the validity of Spanish items or the equivalence of Spanish and English versions (Kaufman & Kaufman, 1990, 2004; Sattler, 2001). As a result, clinicians cannot make informed decisions about test interpretation and remain at a loss when deciding which language is the most appropriate for testing bilingual clients (Hambleton, 2001).

Other available Spanish language IQ measures that are available also suffer from a lack of validation research with appropriate normative samples of Spanish-speaking individuals. For example, the Spanish language version of the WAIS-III, known as the Escala de Inteligencia de Weschler para Adultos – Tercera Edicion (EIWA-III; Weschler, 2008) is commercially available in the United States. The EIWA-III includes the same subtests and constructs as the WAIS-III and is published by Pearson, the same company that publishes the English language WAIS-III. This measure was developed and tested in Puerto Rico to ensure that items were culturally appropriate for Puerto Rican individuals speaking Spanish. With this population, the EIWA-III demonstrates mostly high internal consistencies with mean alpha coefficients ranging
from .73 to .92 and mean standard error ranging .94 to 1.56 for subtests across all age
groups (Pons, et al., 2008).

To date, however, there are no published studies on the validity or reliability of
the EIWA-III with other Spanish speaking populations. Additionally, no research
compares its psychometric properties to the English language WAIS tests. If the EIWA-
III is used for persons outside of Puerto Rico, this lack of psychometric validation and
norms goes against two ITC standards, as well as the standards for educational and
psychological testing which require that psychologists and other professionals refrain
from using a translated version until the reliability and validity of that new measure has
been established for each population with which it is used (AERA, APA, NCME, 1999;
Hambleton, 2001). The danger in administering tests that have not been validated is
that clinicians interpret the results based on an assumption that the test continues to
function in the intended manner (Fantoni-Salvador, 1997). Such assumptions
effectively force minority individuals into inappropriate interpretative categories, thereby
creating a substantial possibility for misdiagnosis and misinterpretation of test results
(Dana, 1993; Todd, 2004). At a minimum, clinicians must provide caveats while
interpreting assessment data and tailor treatment recommendations to different groups
of minority clients (Correa, & Rogers, 2010).

A small amount of validation research has been conducted on a different Spanish
translation of the WAIS-III entitled the Spanish WAIS-III (TEA Ediciones, 2001), adapted
and published in Spain. Research using a Spanish-speaking monolingual sample from
Spain demonstrates that this version of the Spanish WAIS-III supports the same four-
factor structure as the English WAIS-III (Garcia, Ruiz, & Abad, 2003). However, no
comparisons were carried out to determine the equivalency of the tests. Normative data has only been established via the Spanish-speaking sample from Spain. Using a Spanish-speaking sample of Hispanic Americans in Chicago, Renteria, Li, and Pliskin (2008) have conducted the only published validation study on the TEA edition of the Spanish WAIS-III in the US. Their results found adequate reliability and criterion validity for the TEA Spanish WAIS-III. When used with the Hispanic American sample, Spanish WAIS-III subtests had an average internal consistency reliability that was similar to the averages for the sample from Spain (using the Spanish WAIS-III) as well as the North American English-speaking sample (using the English language WAIS-III). Renteria et al. (2008) also identify various areas of bias within the Spanish WAIS-III. For example, they recommend one subtest (Letter-Number Sequencing) that should be omitted because its inadequate alpha coefficients, which indicate limited construct validity. If this subtest is included in analysis, Renteria et al. cautioned that scoring should be more lenient because the structure of the Latin American alphabet makes this task more difficult in Spanish than in English. Lastly, Renteria et al. (2008) highlight specific areas where test bias exists in favor of Spaniards, but lower scores are seen for Spanish-speaking individuals from other Latin American cultures.

In summary, several options are available for Spanish language intelligence testing, each with its strengths and weaknesses. An attractive quality of the K-BIT2 and KABC2 is that both the Spanish and English versions are included in the same test booklets, eliminating the need for evaluators to purchase two separate testing kits. A considerable drawback, however, is the absence of validation data for their Spanish versions. The EIWA-III has published validation data for Puerto Rican populations,
however, its effectiveness with US populations has yet to be tested. Of the three Spanish language measures available, the most researched measure might be the least accessible to mental health professionals in the United States. The Spanish WAIS-III, published in Europe, is the sole measure with validation data available for US populations and the only measure for which specific areas of potential test bias are identified in the research. Clinicians must weigh the pros and cons of each measure in choosing the most appropriate test for their clients.

Culturally Specific Response Patterns Which Affect Validity Scale Scores for Hispanic Americans

Culturally specific response patterns also emerge in the realm of diagnostic measures for psychopathology. For Hispanic Americans, consistent patterns of score elevations are not frequently evident on the clinical scales of multiscale inventories. Instead, patterns are often apparent on validity scales, particularly scales related to minimization of symptoms (Molina & Franco, 1986).

The construct of machismo is among the response patterns that can significantly impact a patient’s self-report measures. Machismo is a gender schema consisting of behaviors, attitudes, and beliefs often espoused by Hispanic American men (Casas, Wagenheim, Banchero, & Mendoza-Romero, 1995). Factors of machismo contain positive aspects related to chivalry and negative aspects related to chauvinism. There is little research in this area, to date, but studies examining machismo, gender roles, and mental health have found that higher levels of machismo and restrictive emotionality can be associated with higher levels of depression and stress among Hispanic American men (Fragoso & Kashubeck, 2000). Therefore, machismo bolsters
the theory that low symptom endorsement does not necessarily indicate subjective well-being among Hispanic Americans. Rather than indicate an absence of symptoms, under-reporting on assessment measures may be more reflective of a general hesitation to disclose symptoms of psychological distress for this clinical population (Correa & Rogers, 2010).

Besides machismo, the conceptualization of extreme response style suggests that individuals of Hispanic and Mediterranean cultures have a tendency to respond either very low or very high when given choices on Likert-type scales in the United States (Hui & Triandis, 1989). It is believed that these individuals consider extreme responses to be more sincere than “less decisive” responses located in the middle of a Likert-type scale. The distinction is most evident for individuals from Hispanic and Mediterranean cultures when contrasted with those from Asian cultures, who tend to respond in the middle of the scale (Zax & Takahashi, 1967). Notably, the language of a test can magnify this cultural response style. In a study that administered the same items in two different languages to bilingual individuals, Gibbons, Zellner, and Rudek (1999) found that participants used more extreme ratings when responding in Spanish than in English. The theory of extreme response style suggests the possibility that Hispanic Americans may be just as likely to over-report symptoms on a measure as they are to under-report. More research is needed in this area.

Validity of Assessment Measures for Ethnic Minority Individuals

Validity of assessment measures used with ethnic minority populations, can be viewed in terms of the etic and emic qualities of the test (Dana, 1993, 2005; Olmedo,
1981). Etic measures assume “universal” applications to individuals of all different cultural groups. Conversely, emic measures are culture-specific and valid for only the groups for whom they were empirically tested. When persons from other cultures are tested and interpreted via mainstream culture, this practice is referred to as imposed etic tests (Berry 1969, 1988, 1989; Dana 2005; Van de Vijver & Hambleton, 1996). That is, test interpretations are made under the assumption that the test items, scales, and constructs all behave in the same manner, regardless of the client’s demographic characteristics. Without validation studies to establish culturally relevant cut scores and interpretation guidelines (or, conversely, to establish that culturally-specific cut scores are not necessary), test developers imply that European American based cut scores are universally valid and generalize to all cultures. This unfounded assumption made by many test developers forces individuals outside of the dominant culture into the same interpretative categories as European Americans, thereby creating a substantial possibility for misdiagnosis and misinterpretation of test results (Dana, 1993; Graham, 1990; Todd, 2005).

Researchers have long criticized translations of multiscale personality inventories that are being made available to clinicians before sufficient validation studies have been conducted, allowing clinicians interpret the results based on an assumption that test continues to function in the intended manner (Fantoni-Salvador, 1997; Rogers, Flores, Ustad, & Sewell, 1995). Mental health practitioners are often unaware of the culturally-specific limitations of tests and unintentionally impose etic effects on individuals being assessed. For this primary reason, it is imperative to validate a translated measure for a new population and determine interpretive guidelines that are best suited to
individuals who are culturally different from the normative sample, particularly when the language of the test items also changes (Geisinger, 1994; Marin & Marin, 1991). Focusing on this limitation, only tests that have been formally translated into Spanish and subsequently validated should be made available for use in clinical practice (American Psychological Association, 2002; Bersoff, 2004; Hambleton, 2001).

Response Styles

Since the inception of standardized assessment measures that rely on a patient's self-report, researchers have agreed that mental health professionals should always make an attempt to determine truthfulness of responses rather than assume all questions are answered in a candid manner. Thus, assessing a client's honesty and forthrightness is a vital part of an evaluation (Hathaway & McKinley, 1940). Many standardized and widely used assessment measures, such as the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and the Personality Assessment Inventory (PAI; Morey, 1991, 2007) contain validity scales to gauge response styles in an effort to determine whether an examinee's report should be trusted as accurate. This section discusses response styles and culminates with their specific application to PAI validity scales.

Response styles are a group of empirically established patterns patients can exhibit during the process of answering questions in a psychological assessment. An examinee's test-taking attitudes at the time of evaluation and their particular response patterns can affect the validity of test data obtained in a psychological evaluation with the potential for distorted assessment results (Rogers, 1984; Rogers, 1997; Rogers,
Bagby, & Dickens, 1992). This intentional distortion is especially salient if clients choose to purposely overreport or underreport their symptoms and impairment. Psychological assessment must take into account and incorporate methods for their detection in psychological assessments in order to minimize misdiagnosis of clients (Resnick, 1984; Rogers, 1997; Rogers, 2008; Rogers & Schuman, 2005).

Throughout the history of psychological assessment, several response styles have been thought to influence assessment results. Some response styles reveal people who intentionally under-report negative symptoms and personal qualities. Paulhus (1984) found strong empirical support for a two-component model of socially desirable responding: (a) self-deception, where individuals believe their own false reports, and (b) impression management, when individuals consciously provide disingenuous responses that will make them appear favorable to others. These core facets of simulated adjustment have been studied under different names by various researchers. Whether referred to as “self-deception” and “other-deception” (Sackeim & Gur, 1978), “desirability” and “defensiveness” (Kusyszyn & Jackson, 1968), or using Paulhus’ terms, the implication of these response styles is that the authenticity of information gleaned from self-reports stands at the mercy of patients’ own misinformed versions and intentional distortions of their clinical conditions.

Disingenuous responding such as symptom minimization can be done unintentionally by the patient, as in self-deception. The false reports can also be purposeful, however, as in impression management and other-deception (Kusyszyn & Jackson, 1968; Paulhus, 1984; Paulhus, Bruce, & Trapnell, 1995; Sackeim & Gur, 1978; Whyte, Fox, & Coxell, 2006). This distinction parallels the non-intentional feigning of
somatization disorders and the deliberate fabrication of symptoms found in factitious disorders and malingering (DSM-IV-TR; APA, 2000). For both response styles, the chief distinguishing factor involves whether the client is purposely reporting false symptoms.

Rogers (1984) expanded the conceptualization of response styles to encompass four basic styles, described in Table 1.

Table 1

Description of Response Styles

<table>
<thead>
<tr>
<th>Response Style</th>
<th>Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliable</td>
<td>Individuals with this approach to a psychological evaluation generally attempt to answer assessment questions honestly.</td>
</tr>
<tr>
<td>Irrelevant</td>
<td>Individuals are haphazard or inconsistent with their responses to test items.</td>
</tr>
<tr>
<td>Defensive</td>
<td>Individuals deny or minimize symptoms of psychological impairment.</td>
</tr>
<tr>
<td>Malingering</td>
<td>Individuals purposely falsify or exaggerate symptoms for an external objective.</td>
</tr>
</tbody>
</table>

Defensiveness and malingering are two response styles that share elements of dissimulation motivated by external goals. Both response styles can cause significant concern for mental health systems. For example, underreporting of symptoms is a chief clinical concern because individuals engaging in this response style appear less impaired than they actually are, and could consequently avoid necessary psychological intervention (Meehl & Hathaway, 1946; Rogers & Shuman, 2005). Conversely, individuals reporting exaggerated or false symptoms of mental disorders might inappropriately use resources intended for individuals in genuine need of them. A more
detailed analysis of defensiveness and malingering is addressed in the next three sections. Important cultural issues and each response style’s effect on psychological assessment results is also be addressed.

Malingering

Individuals who purposely exaggerate their condition or report false symptoms are generally thought to fall into two main categories: factitious disorders and malingering (Overholser, 1990). Patients diagnosed with factitious disorders fabricate symptoms unmotivated by external rewards. Instead, their motivation to feign is an internal drive, producing personal and intangible benefits (Gorman, 1982; Hagglund, 2009). The DSM-IV-TR narrows this conceptualization by specifying that the person’s motivation for symptom fabrication must be to assume “the sick role” and garner the attention that comes with being treated as a patient (APA, 2000).

When patients intentionally report false or grossly exaggerated symptoms (i.e., feigning), this presentation can have significant consequences for diagnosis and subsequent clinical interventions. In a clinical setting, even ambiguous evidence of feigning can prevent prospective patients from receiving mental health services (Rogers, 1997, 2008) because in settings where resources are scarce, many mental health professionals believe it is their responsibility to ensure that only the truly sick are given the limited availability of mental health treatment (Resnick, 1984). In forensic settings, the ramifications of suspected feigning can be even more serious. Not only might individuals be denied mental health care, but the classification of malingering can be used to entirely discredit their clinical presentations at all stages of the trial process.
(Rogers & Shuman, 2005). For example, the criminal justice system attempts to ensure that only defendants with severe disorders, not feigners, are excused from culpability and punishment in a verdict of Not Guilty by Reason of Insanity. Classifications of malingering can damage future treatment because it is often difficult for them to prove the genuineness of their disorders, once categorized as malingerers. For this reason, a thorough assessment must be conducted before making such a consequential classification (Berry, Baer, Rinaldo, & Wetter, 2002).

As mentioned earlier, the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR; APA, 2000) identifies malingering as the deliberate production of feigned symptoms by a person seeking some form of external gain. Other DSM diagnoses, such as factitious disorders and somatoform disorders, also involve the production of false symptoms, but the key difference is the underlying motivation. According to the American Psychiatric Association (2000), only malingerers intentionally falsify symptoms for the purpose of obtaining an obvious external benefit, such as financial compensation, exemption from duty, or leniency from the criminal justice system. However, malingering can be difficult to detect accurately because an individual’s method of feigning can vary substantially from client to client (Reid, 2000). Furthermore, some researchers specify that malingering, per se, cannot be detected by any psychological measures because these tests cannot identify a person’s often multi-dimensional motivation. Thus, assessment measures can only evaluate feigning. Motivation must be extrapolated from additional data such as clinical interviews, observations, and collateral sources (DeClue, 2002; Rogers, 1997).
Criticisms of the *DSM-IV-TR* definition and disagreement among researchers further complicate professionals’ ability to accurately classify malingering (DeClue, 2002). Discrepancies in the field can lead to confusion regarding important points of focus during a comprehensive assessment. For instance, the broad *DSM* definition stated above is generally accepted, but experts in malingering often disagree about the validity of its operationalization including its focus on screening indicators made by the APA’s diagnostic manual (Rogers, 2008). These indicators, outlined by the *DSM-IV* are presented (American Psychiatric Association, 2000, p. 739):

Malingering should be strongly suspected if any combination of the following is noted:

1. Medico-legal context of presentation (e.g., the person is referred by an attorney to the clinician for examination)
2. Marked discrepancy between the person’s claimed stress or disability and the objective findings
3. Lack of cooperation during the diagnostic evaluation and in complying with the prescribed treatment regimen
4. The presence of antisocial personality disorder

Some professionals advocate that the *DSM-IV* indices provide good guidelines for identifying potential malingerers during an assessment and even suggest broadening the concept of malingering to include responses that distort an honest portrayal of symptoms in any manner (Meyer & Deitsch, 1996). Other professionals, however, contest this viewpoint, (Rogers & Shuman, 2005). Rogers (1997) is sharply critical of *DSM-IV*’s approach, citing data from a study that found the *DSM-IV* screening indicators misclassified nearly four genuine patients (resulting in false positives) for every malingerer that was correctly identified. In fact, the *DSM-IV* indicators accurately
identified malingerers only 20.1% of the time (Rogers, 1990). Because of the serious consequences inherent in an erroneous classification of malingering, many researchers contend that the false positive rate encountered through using DSM-IV indicators are clearly not acceptable (Berry, Baer, Rinaldo, & Wetter, 2002; DeClue, 2002; Rogers & Shuman, 2005).

Scholars and mental health professionals (Cunningham & Reidy, 1999; Hare, 2003) argued cogently that DSM-IV guidelines are inadequate because most examinees undergoing forensic evaluations will meet several of the screening indicators, even if they are not malingering, simply due to the nature of the assessment. Specifically, all defendants will meet the first indicator (i.e., medico-legal context). It is likely that the majority of criminal defendants will also qualify for the fourth indicator because many offenders meet criteria for Antisocial Personality Disorder. Thus, many criminal forensic patients meet two indicators in the DSM-IV purely by default. Such research findings should prompt professionals to apply DSM-IV indices very cautiously. It is, perhaps, most advisable to treat them only as screening indicators and use them to prompt a more thorough evaluation.

In contrast to Meyer and Deitsch’s (1996) suggestion to broaden the concept of malingering, Rogers (1997) proposes narrower definitions. Specifically, the classification of malingering is reserved solely for cases where there is definite evidence of deliberate exaggeration or fabrication of psychological problems. Malingering is a conscious choice, motivated for external gain. Thus, Rogers’ approach is more conservative in classifying examinees as malingerers, emphasizing that until clear evidence of motivation is established, examinees should only be referred to as
“feigners.” This more conservative approach is focused on minimizing false positives (DeClue, 2002; Rogers et al., 1992) and emphasizes the practitioner’s intent on minimizing the risk of misclassifying individuals as malingers (Melton, Petrila, Poythress, & Slobogin, 1997).

Finally, some professionals recommend adherence to guidelines or a specified model for the evaluation of malingering, particularly in situations where assessment findings are likely to be presented in court; expert evidence should be standardized with demonstrable scientific rigor (DeClue, 2002; Meyer & Deitsch, 1996). Meyer and Deitsch (1996) provide a checklist for malingering, which gives some guidance for clinical decision making. However, no empirical support exists regarding the reliability or validity of this checklist. Additionally, their interpretive guidelines utilize their aforementioned broad conceptualization of malingering (DeClue, 2002). For practitioners wishing to espouse a more stringent conceptualization, Rogers (1997) presents two models for malingering assessment: (a) a threshold model for clinicians to decide when they should evaluate feigning more thoroughly and (b) a clinical decision model which requires additional sources of data so that no single measure is solely relied on for classification of malingering. Using the Structured Interview of Reported Symptoms (SIRS; Rogers et al., 1992), the threshold model for suspected malingering is based on (a) four or fewer SIRS scales in the honest range, or (b) one to two SIRS scales in the probable range. By contrast, the clinical decision model utilizes (a) one or more scales in the definite feigning range or (b) three or more scales in the probable feigning range. This model leads to accurate classification of more than 90% of individuals undergoing evaluations (Rogers, 1997). Rogers’ models provide a clear,
theoretically sound and empirically supported framework for clinicians to interpret findings and describe the degree of certainty about whether a subject is feigning. Therefore, Rogers’ models may be more useful for practicing clinicians than the general guidelines provided by Meyer and Deitsch and the *DSM-IV* (DeClue, 2002).

Models of clinical decision making for the classification of malingering fall into two general categories: hypothesis-testing models and a linear best-fit models (Rogers & Shuman, 2000). Examiners using a hypothesis-testing model first formulate a working hypothesis about the patient’s diagnosis or classification (e.g., malingering) toward the beginning of the evaluation, and proceed to gather data that confirms or disconfirms their hypotheses. If a hypothesis is disconfirmed, a new hypothesis is subsequently formed and tested. In a linear best-fit model, the examiner conducts the assessment in two phases. The first phase consists entirely of data collection. The examiner gathers comprehensive data, and refrains from formulating interpretations that could bias the assessment. In the second phase, the examiner compares competing hypotheses and forms opinions and conclusions based on the relative strengths of each hypothesis.

Borum, Otto, and Golding (1993) address potential problems with hypothesis-testing approaches and recommend that experts always test alternative hypotheses to prevent issues of “cherry-picking” only the data that supports an initial hypothesis. Although the hypothesis-testing model may be most often used by forensic examiners, Rogers and Shuman (2000) also advocate using the linear best-fit model to test alternative hypotheses in malingering evaluations and minimize issues such as: primacy bias, confirmatory bias, and over-reliance on unique data.
Defensiveness

The second response style, critically important to the current study, is defensiveness. As previously noted, defensiveness during a psychological assessment is apparent when examinees attempt to downplay their symptoms of psychological impairment (Rogers, 1984). In many cases, defensive response styles emerge in distinct patterns among members of ethnic minority populations. For example, in a classic study, Molina and Franco (1986) found significant differences in self-disclosure based on ethnicity and gender in non-clinical populations. Overall, Mexican Americans tended to self-disclose less than their European American counterparts. Moreover, Mexican American men self-disclosed even less than Mexican American women. If these findings hold true for clinical populations, it is imperative that clinicians remain aware of unique cultural response patterns as part of a thorough assessment. If individuals from a different cultural background, such as Latino, appear to respond in a guarded or defensive manner during psychological assessments, this presentation can have a significant impact on the validity of their clinical profiles and the subsequent accuracy of their diagnoses (Helms, 1992). Specific cultural issues as they relate to response styles on standardized assessment measures will be addressed later.

Notably, the constructs of defensiveness and social desirability are often used somewhat interchangeably in assessment literature. For example, Greene (2008) points to a meta-analysis of MMPI defensiveness measures by Baer, Wetter, and Berry (1992), which shows that the largest effect sizes for defensiveness were found on a measure of social desirability, specifically, the Wiggin’s Social Desirability Scale (Sd). Part of the overlap between constructs may be due to the structure of so-called
defensiveness scales. While defensiveness scales focus on minimized or denied psychological impairment and patient characteristics (e.g., the MMPI K scale), other scales focus on general dishonesty or social desirability (Rogers, 2008).

The Development of Detection Strategies for Malingering and Defensiveness

Detection strategies are standardized, theoretically based methods which have been empirically tested and validated for differentiating between specific response styles used in standardized assessment measures (Rogers, 1997). Detection strategies for malingering can be divided into two main categories: unlikely and amplified. Unlikely detection strategies focus on the endorsement of highly unusual or “bogus” symptoms to determine feigning. Amplified detection strategies focus on the intensity of reported symptoms and determine whether it is much greater than typically reported by genuine patients (Rogers & Correa, 2008).

In 1997, Rogers described a number of detection strategies for feigned psychopathology. Table 2 briefly describes each strategy and classifies them into the two broad domains. In understanding the application of these detection strategies, Miller’s work (2001) provides a useful illustration in creating a malingering screen, the Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001). The M-FAST included scales to assess the following detection strategies in her measure: reported vs. observed (RO), extreme symptomatology (ES), rare combinations (RC), unusual hallucinations (UH), unusual symptom course (USC), negative image (NI), and suggestibility (S).
## Table 2

**Description of Detection Strategies for Malingering**

<table>
<thead>
<tr>
<th>Detection Strategy</th>
<th>Domain</th>
<th>Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rare symptoms</td>
<td>Unlikely</td>
<td>Focuses on symptoms that rarely occur in psychiatric patients; over-endorsement of uncommon symptoms may indicate that the client is exaggerating or feigning.</td>
</tr>
<tr>
<td>Improbable symptoms</td>
<td>Unlikely</td>
<td>Focuses on the number of symptoms endorsed by a person which are so outlandish, that they are highly unlikely to be true symptoms of a disorder. The presence of multiple improbable symptoms are often associated with feigning.</td>
</tr>
<tr>
<td>Symptom combinations</td>
<td>Unlikely</td>
<td>Focuses on inquiries about true psychological symptoms. However, some unusual symptom pairs are rarely observed in genuine patients. Over-endorsement of unusual combinations may indicate malingering.</td>
</tr>
<tr>
<td>Reported vs. observed symptoms</td>
<td>Unlikely</td>
<td>Focuses on the clinician's own observations compared to the symptoms that the client reports. When the client reports a much higher number of observable symptoms, it may be because the person is reporting false symptoms.</td>
</tr>
<tr>
<td>Spurious patterns</td>
<td>Unlikely</td>
<td>Focuses on patterns of response that are characteristic of malingering, but are very uncommon in clinical populations.</td>
</tr>
<tr>
<td>Erroneous stereotypes</td>
<td>Unlikely</td>
<td>Focuses on whether the person being evaluated reports an excessive number of misconceptions about mental disorders held by the general population. If so, the issue of feigning is raised, as people who do not actually suffer from a particular disorder may be misinformed about symptoms and their presentation.</td>
</tr>
<tr>
<td>Obvious symptoms</td>
<td>Amplified</td>
<td>Focuses on whether the person being evaluated reports a larger-than-expected number of symptoms that are clear indicators of psychopathology.</td>
</tr>
<tr>
<td>Subtle symptoms</td>
<td>Amplified</td>
<td>Focuses on whether the person endorses a very large number of symptoms seen as common difficulties not necessarily indicative of mental disorders.</td>
</tr>
<tr>
<td>Symptom selectivity</td>
<td>Amplified</td>
<td>Focuses on how selective examinees are in their endorsement of psychological problems. Malingerers tend to endorse a wider array of symptoms from various disorders than genuine patients typically do.</td>
</tr>
<tr>
<td>Symptom severity</td>
<td>Amplified</td>
<td>Focuses on how the person being evaluated characterizes the intensity of their symptoms. Genuine patients will typically identify some of their symptoms as being worse than others. However, malingerers tend claim that many of their symptoms are “extreme.”</td>
</tr>
</tbody>
</table>
Four scales utilize similar detection strategies to those identified by Rogers (1997). They include rare symptoms (UH), symptom combinations (RC), reported vs. observed (RO), and severity of symptoms (ES). These strategies rely on unlikely presentation of symptoms (Vitacco, Jackson, Rogers, Neumann, Miller, & Gabel, 2008). Combining these detection strategies, research generally finds that the M-FAST is a valid screen for the detection of feigned psychopathology (Guy, Kwartner, & Miller, 2006; Jackson, Rogers, & Sewell, 2005).

Detection strategies for defensiveness can also be generally classified into distinct categories: idealized attributes and denial of impairment. Table 3 describes each strategy and classifies them into the broad domains.

Table 3

<table>
<thead>
<tr>
<th>Detection Strategy</th>
<th>Domain</th>
<th>Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social desirability</td>
<td>Idealized attributes</td>
<td>Focuses on individuals who attempt to create a very favorable image and potentially identifies them as persons who are denying maladjustment.</td>
</tr>
<tr>
<td>Denial of personal faults</td>
<td>Denial of impairment</td>
<td>Focuses on the idea that people who are minimizing maladjustment will also deny personal shortcomings and negative behaviors.</td>
</tr>
<tr>
<td>Denial of patient characteristics</td>
<td>Denial of impairment</td>
<td>Focuses on attributes that are commonly endorsed by clinical populations and considers lack of endorsement as a sign of defensiveness.</td>
</tr>
<tr>
<td>Blended Strategy</td>
<td>Both</td>
<td>Focuses on a combination of endorsing overly positive attributes and denying common shortcomings.</td>
</tr>
<tr>
<td>Spurious patterns of simulated adjustment</td>
<td>Both</td>
<td>Focuses on scale configurations that are frequently seen in defensive individuals, but not commonly found in clinical and community samples.</td>
</tr>
</tbody>
</table>
Detection strategies for defensiveness are considered to be less sophisticated than those for malingering (Rogers, 2008). Strategies for defensiveness have several limitations including imprecision in operationalizing strategies and overlap. Overlap is evidenced by (a) blended strategies, and (b) attempts to infer defensiveness from endorsement of overly positive traits (e.g., social desirability strategy).

Assessment of Malingering and Defensiveness

Feigning is notoriously difficult to detect by clinical interview alone and even experienced mental health professionals are often unsuccessful. Early research (Bourg, Connor, & Landis, 1995) reveals that clinicians conducting interviews of examinees are generally poor evaluators of malingering. This lack of success is likely due to the fact that clinical interviews are not standardized and rely almost exclusively on the mental health professional’s own judgment (Borum, Otto, & Golding 1993; DeClue, 2002; Geller et al, 1990; Meagher, 1919; Pope, 1919; Resnick, 1984). When clinicians do not perceive the client’s deceptive intent, or when they do not make sufficient inquiries, feigning can go undetected (Rogers, 1997; Rogers & Shuman, 2005). Thus, valid measures of feigning are crucial. A structured interview like the Structured Interview of Reported Symptoms (SIRS; Rogers et al., 1992), and its recently published second edition, the SIRS-2 (Rogers, Sewell, & Gillard, 2010), are comprehensive measures designed to evaluate feigned mental disorders and are widely considered the gold standard for the detection of feigned mental disorders (Blau, 1998; DeClue, 2002; Lally, 2003; Rogers, 2001; Rogers, 2008). Especially in forensic contexts, the SIRS is the most researched specialized measure for the assessment of
feigning. However, the current study focuses primarily on the use of self-report measures. Therefore, the following section addresses psychological assessment of response styles using such measures.

Assessment of Response Styles Using Multiscale Inventories

Guy et al. (2006) observed that a major advantage of multiscale inventories is their application of embedded validity scales for the assessment of response styles. The advent of multiscale inventories first allowed for the evaluation of feigners and honest responders through systematically comparing differences between the two criterion groups. The original MMPI (Hathaway & McKinley, 1940) fundamentally changed the assessment of response styles and malingering. According to Meehl and Hathaway (1946), clinicians must assume patients could be motivated to deliberately alter their symptom presentation. These early researchers found it important to include scales to assess response styles in order to determine the genuineness of a client’s self-report. What follows is a brief overview of the MMPI-2 and a more in-depth discussion of the PAI, which is the primary focus of the current study.

The MMPI-2

The Minnesota Multiphasic Personality Inventory 2 (MMPI-2; Butcher et al., 1989) is a widely researched multiscale inventory that includes well-established validity scales. Basic MMPI-2 validity scales are designed to determine whether examinees are responding in an inconsistent manner, defensive manner (underreporting), or feigning (over-reporting) symptoms of severe psychopathology (Greene, 2000). Detection
strategies employed by the MMPI-2 validity scales allow researchers to caution practitioners against relying exclusively on certain scale elevations. When considering the MMPI-2, it is important to distinguish between the two detection strategies used in the F scale family (e.g., F, Fb, Fp, Fptsd). The Infrequency Psychopathology (Fp) scale uses rare symptoms, whereas the Infrequency (F) and Infrequency Back (Fb) scales use quasi-rare symptoms strategy. Unlike rare symptoms, quasi-rare symptoms are those which are found very infrequently in the general population, but not necessarily in clinical populations where the MMPI-2 is frequently used.

Some researchers conclude that the MMPI-2 F scales are generally deemed effective in identifying overreported psychopathology (Sellbom & Bagby, 2010). Others criticize the use of quasi-rare symptoms for the detection of feigning, stressing that some symptoms (e.g., hallucinations) are rare in community samples but common in some clinical populations. Therefore, the endorsement of such symptoms should not necessarily be equated with malingering. For example, patients with genuine psychotic disorders are often show elevation on scales using quasi-rare symptoms and may be miscategorized as malingerers (Gough, 1947; Rogers & Bender, 2011; Rogers, Sewell, Martin, & Vitacco; 2003).

The MMPI-2 was originally designed with two scales to assess under-reporting: the Lie (L) scale and the Correction (K) scale (Greene, 2000). The L scale was designed to identify individuals attempting to present themselves in an overly positive light. It is primarily associated with individuals who are denying minor faults and has been labeled as a social desirability scale (Cloak, Kirklen, Strozier, & Reed, 1997).
contrast, the K scale could indicate defensiveness or a patient’s lack of insight regarding their symptoms of psychopathology (Greene, 2000).

Further emphasizing the overlap between detection strategies for under-reporting employed by the L scale, Burish (1976) investigated the construct validity of the scale as a measure of defensiveness. Findings from his study suggest that, while the L scale has been identified as a measure of social desirability, individuals with high L scores use defensive maneuvers to cope during stressful situations. The L scale was also found to correlate significantly with the Denial scale (another measure of defensiveness), but did not correlate with nondefensive MMPI-2 scales (Burish, 1976). Overlap between some scales and not others may be indicative of poor construct validity in measures of underreporting.

A meta-analysis by Baer and Miller (2002) suggests that L and K scales are reasonably accurate in detecting uncoached underreporters, because the L scale shows the highest specificity. However, the researchers specify that detection of coached feigners and the incremental validity of MMPI-2 supplementary scales require further investigation. Additionally, they call for a clear distinction to be made regarding underreporting in different contexts. Specifically, Baer and Miller (2002) emphasize the difference between those who respond defensively primarily due to situational demands (i.e., child custody cases and personnel selection) and those who are concealing psychopathology. Because underreporting remains more difficult to detect than overreporting, suspicions about underreporting that are triggered by elevations on these scales should be investigated through interview, behavioral observations, other self-
report inventories, and collateral sources of information, as available and appropriate (Berry et al., 2002).

Cloak, Kirklen, Strozier, and Reed (1997) call for a new way of conceptualizing MMPI-2 validity scales with clearer definitions and consistency among terms. Their factor analysis of validity scale item responses from MMPI data that examined L, F, and K scores. Analyses yielded 4 major factors: Minimizing, Exaggerating, Cynicism, and Psychological Distress. The authors suggest the Minimizing and Exaggerating factors seem to confirm the utility of scales measuring social desirability, defensiveness, faking, and malingering, but also suggest that their inferences point to a need for response bias scales with more distinct definitions and greater internal consistency (Cloak, et al., 1997).

The recently published MMPI-2 Restructured Form (MMPI-2-RF; Tellegen & Ben-Porath, 2008) includes eight validity scales, including some major revisions of the original MMPI-2 validity scales, plus the addition of one new scale—the Infrequent Somatic Responses (Fs) scale. Recently, Sellbom, Toomey, Wygant, Kucharski, and Duncan (2010) examined the utility of the MMPI-2 RF validity scales within a criminal forensic sample. Using the SIRS as the criterion, they found that MMPI-2-RF validity scales were able to adequately differentiate between overreporting and genuine responding defendants. The F-r and Fp-r scales performed the best in differentiating between the two groups, with very large effect sizes (Cohen's $d$) of 2.11 and 2.07, respectively. A second outpatient study confirmed that Fp-r best differentiated between simulation groups and genuine patients (Sellbom & Bagby, 2010). The few studies on
malingering and the MMPI-2-RF published to date, provide positive findings for the Fp-r scale and feigned mental disorders.

The PAI

The Personality Assessment Inventory (PAI; Morey, 1991) is a second-generation multiscale personality measure that also uses validity scales to identify response styles including malingering and defensiveness. Several studies have found exceptionally strong support for all three validity scales in differentiating feigners from honest responders, with Cohen’s $d$’s greater than 2.00 for each scale (Fernandez et al., 2008; Morey & Lanier, 1998). Other studies have found that some scales clearly perform better than others for detecting overreporting (Bagby, Nicholson, Bacchiochi, Ryder, & Bury, 2002; Boccaccini, Murrie, & Duncan, 2006) or that no scale is especially effective for this purpose (Calhoun, Earnst, Tucker, Kirby, & Beckham, 2000; Edens, Poythress, & Watkins-Clay, 2007). Efficacy of scales can vary widely depending on research design and characteristics of the sample. For this reason, it is especially important to understand how scales perform for groups of individuals with characteristics similar to those of the examinee in question.

Of the three PAI validity scales, Negative Impression Management (NIM) is most often used to assess malingering. NIM uses a rare-symptoms detection strategy, and its items were selected because of their low level of endorsement among clinical and non-clinical samples (Morey, 2007; Rogers, et al., 2011). Although other interpretations must be considered, high NIM scores may indicate examinees are exaggerating symptoms or endorsing a large amount of extremely bizarre symptoms. For instance,
Boccaccini et al. (2006) found that the NIM scale items ($d = 1.54$), outperformed other PAI feigning indices such as RDF ($d = 0.21$) for the detection of feigning. One major advantage in using the PAI is that, unlike the MMPI-2, NIM items do not overlap with validity or clinical scales. Thus, the PAI does not suffer from the same problems as the MMPI-2’s F scale and Fb whose atypical items in non-clinical populations may misclassify honest but impaired responders. However, Rogers and Bender (2003) caution that the PAI should not be used as the sole measure used to detect malingering because only extreme elevations on NIM and MAL are indicative of feigning. Morey (2007), himself, cautions that the NIM scale is “not a malingering scale per se” (p. 29) as exaggerated presentation and endorsement of unlikely symptoms may be a prominent component of many Axis I and Axis II disorders. In these cases, high NIM scores do not indicate malingering, but render a profile uninterpretable.

In their study of inpatients, Rogers, Gillard, Wooley, and Ross (2011) found the NIM scale was routinely elevated in patients with extensive trauma histories ($M = 71.96$) and especially elevated among patients with PTSD and dissociative identity issues ($M = 85.85$). This finding raises the possibility that NIM items are experienced more frequently among patients with specific types of severe pathology. Although NIM appeared affected by trauma, especially when dissociative symptoms were prevalent, other feigning indicators (i.e. MAL and RDF) using the more complex spurious patterns detection strategy did not demonstrate such elevations. These findings suggest the possibility that the increased complexity of unlikely detection strategies (e.g., from rare symptoms to spurious patterns) may improve the classification accuracy for certain patient samples.
Sellbom and Bagby’s review (2008) found that NIM and MAL proved effective using both known-groups and simulation designs. The Malingering Index (MAL), which uses a *spurious patterns* detection strategy to examine different response configurations indicative of feigned mental disorders is designed to be used with NIM scale scores to provide a more specific indicator of malingering (Morey, 2007; Selbom & Bagby, 2008). Hawes and Boccaccini (2009) conducted a meta-analysis of the PAI and feigning. Simulation studies found that each validity measure is a strong predictor of uncoached (NIM, \(d = 1.48\); MAL, \(d = 1.15\); RDF, \(d = 1.13\)) and coached malingering (NIM, \(d = 1.59\); MAL, \(d = 1.00\); RDF, \(d = 1.65\)). When feigners were compared to unimpaired honest responders as opposed to patients, cut scores of NIM and MAL resulted in the highest overall classification rates for identifying feigning. These results apply only to simulation research, however. Specifically, NIM effect sizes for studies with patient comparison groups and known-groups comparisons were not significantly different. Additionally, the difference between MAL effects from patient comparison group studies and known-groups studies was not statistically significant.

Rogers’ Discriminant Function (RDF; Rogers, Sewell, Morey, & Ustad, 1996), tends to show much more variability between studies than other PAI indicators of feigning. The Rogers’ RDF is a statistically derived discriminant function that uses the weighted combinations of 20 PAI scores. It was developed to distinguish the PAI profiles of genuine patients from those who are simulating specific diagnoses. Research on the effectiveness of RDF is mixed, with some studies reporting large effect sizes (Morey, 2007) and others reporting, very low effect sizes for RDF (Sellbom & Bagby, 2008). The large discrepancy in effect sizes appears to be related to differences
in study design and sample characteristics. Specifically, RDF proved to be more effective than NIM and MAL in simulation studies but failed to distinguish between feigners and honest responders in known-group comparisons (see Sellbom & Bagby, 2008). This disparity for RDF could also be attributed to differences in setting (i.e., forensic vs. clinical; Hawes & Boccaccini, 2009). Thus, the discrepancy in research findings may be due to experimental design and samples.

The three indicators of underreporting on the PAI are: PIM, DEF, and CDF. The Positive Impression Management (PIM) scale is designed to detect individuals who are denying negative attitudes, behaviors, or traits. It includes items that are frequently endorsed in both clinical and non-clinical samples (Morey, 2003). Therefore, persons who exhibit elevations on the PIM scale are thought to be presenting themselves in an overly positive manner, thereby responding to test items in a dishonest style that masks negative attributes. In addition to PIM, two more PAI indexes assess defensiveness. The Defensiveness Index (DEF) examines eight different configural patterns that are frequently observed among individuals attempting to present themselves in the best way possible. In contrast, the Cashel Discriminant Function (CDF) uses the scores of six different PAI scales to create a function score. It was derived from a study that asked participants to present themselves favorably, while stressing that their self-report should be convincing (Cashel, Rogers, Sewell, & Martin-Cannici, 1995). The Cashel CDF has been found to be more accurate than either the PIM or DEF scores in detecting defensiveness among samples of male inmates and male undergraduate students (Cashel et al., 1995; Morey, 2007).
Research suggests the constructs defining the PAI defensiveness are unclear, limiting their effectiveness, and requiring further refinement (Rogers, 1988). Problems with accurately identifying under-reporters on the PAI arise from the same issues of construct validity noted by Baer and Miller (2002) regarding defensiveness and social desirability on the MMPI-2. In a study using university students, Peebles and Moore (1998) suggest PIM and DEF adequately identify socially desirable responding on the PAI. However, they also determined that a lower PIM cut score (raw score \( \geq 18 \)) was more effective in correctly classifying defensive responders than the cut score suggested in the PAI manual (raw score \( \geq 23 \)). The seemingly interchangeable use of the two terms indicates there remains a considerable level of overlap between the constructs of socially desirable responding and defensiveness.

Due to the modest sensitivity of defensiveness indicators, Cashel, et al. (1995) also recommend using a lower cut score than suggested in the PAI manual in order to accurately identify under-reporters. In their study using male inmates and male college students, Cashel et al. (1995) found Morey’s recommended PIM cut score of \( \geq 68T \), misclassified approximately 5 out of every six defensive profiles. They proposed a cut score of \( \geq 57T \) to increase sensitivity. Using this new cut score as a benchmark, nearly half of the defensive profiles were accurately classified. However, they note the PIM scale combines defensiveness and socially desirable responding, so the construct of defensiveness may need further refinement. Additionally, characteristics of defensiveness are both situation and population specific (Rogers, 1988).
The Bipolarity Hypothesis

As noted previously, malingering and defensiveness are response styles that share deliberate efforts at distorting clinical characteristics from under to over-reporting (Rogers, 1984). Therefore, these two styles are often considered to be two “endpoints on a continuum” particularly for multiscale inventories, such as the MMPI-2 (see Greene, 2008, p. 167) and PAI (Morey, 2007). For example, the MMPI-2’s F-K Index, originally named the Gough Dissimulation Index (Gough, 1950), assesses the relationship between F and K scales in order to determine both feigning and defensiveness on MMPI protocols. In support of the bipolarity hypothesis, high scores on this index are indicative of feigning, whereas low scores indicate defensiveness. Similarly, two primary validity scales on the PAI, NIM (feigning) and PIM (defensiveness) display a low to moderate inverse relationship (Morey, 2007). This negative correlation partially corroborates the bipolarity hypothesis, indicating that feigners tend to have low scores on measures of defensiveness and vice versa. In addition, Morey and Lanier (1998) provide further evidence for the bipolarity hypothesis in a PAI meta-analysis. Their results found that scores on the PAI defensiveness indicators PIM and DEF are positively correlated with each other and negatively correlated with the three PAI measures of feigning (i.e., NIM, MAL, and RDF).

In support of the bipolarity hypothesis, several have found that feigners, indeed, exhibit lower scores on measures of defensiveness. For example, Graham, Watts, and Timbrook (1991) found markedly suppressed scores on the MMPI-2’s K scale for both male ($M = 35.8$ T) and female ($M = 32.7$ T) feigners in a simulation design. In an MMPI-2 meta-analysis, Rogers, Sewell, Martin, and Vitacco (2003) also found that most
feigners do not show elevations on K. These findings offer some support for the bipolarity hypothesis. However, Rogers et al. (2003) emphasized that it is yet to be determined whether the absence of defensiveness effectively discriminates feigned from genuine profiles.

Currently, no studies examine the bipolarity hypothesis within different cultural contexts. For example, Hispanic American individuals might tend to have a cultural response style where they are reticent to disclose both personal and potentially negative information within the context of a psychological evaluation (Correa & Rogers, 2010). On this point, Correa (2010) found that approximately one-third of patients instructed to respond honestly on the Spanish SIRS-2 attained elevated scores on the Defensiveness (DS) scale, which measures defensiveness and denial of everyday problems. If honest responders tend to respond defensively, significant negative correlations between measures of defensiveness and malingering may not exist for this cultural group.

Malingering and Defensiveness among Mexican Americans

Multiscale Inventories

Cultural differences appear to affect the perceived openness of Hispanic Americans on multiscale inventories. As previously discussed, Molina and Franco (1986) found in the general population that Mexican Americans tended to self-disclose less than European Americans. Moreover, Mexican American men self-disclosed even less than Mexican American women.
Early MMPI research conducted with Hispanic American individuals corroborates the increased pattern of perceived defensiveness. In an early meta-analysis by Campos (1989), several studies consistently found significantly higher L scale elevations among Hispanic Americans when compared to European Americans across clinical samples. Likewise, L scale elevations have also been found for Hispanic American women on the MMPI-2 (Callahan, 1998). Elevations on that scale typically indicate the examinees are deliberately distorting their presentation in order to present themselves in the best possible light (faking good; Greene, 2000). While the L scale is commonly thought to detect those who are denying minor faults, this response style could also indicate a culturally-specific hesitation to express personal feelings, a sensitivity about stigmatization, and a selectivity in disclosing personal problems.

Current research begins to address how cultural differences may affect the standardized assessment of response styles such as feigning. In studying the clinical utility of the Spanish-language MMPI-A for the detection of feigners, Lucio, Duran, Graham, and Ben-Porath (2002) evaluated clinical and non-clinical Spanish-speaking adolescents in Mexico. They studied four indicators (F, F1, and F2 scales, and F-K index) on the Mexican Spanish translation of the Minnesota Multiphasic Personality Inventory-Adolescent (MMPI-A; Lucio, 1998). They found these indicators effectively discriminated between feigners and honest responders for both groups, with high PPP and NPP values associated with the F scale cut scores for male (PPP = .82; NPP = .89) and female (PPP = .86; NPP = .78) adolescents. While the measure was effective in classifying their particular samples, Lucio et al. (2002) caution against generalizing their
findings to Hispanic adolescents in countries other than Mexico, highlighting cultural differences. In particular, Lucio et al. (2002) suggest that different cut scores might be necessary for the MMPI-A because Hispanic American adolescents in the United States tend to be less forthcoming than their Mexican counterparts.

Mendoza-Newman (2000) also acknowledged the limited generalizability of research findings across different cultural groups and advocated the need for different cut scores to counteract the effects of acculturation and culturally-specific response styles on profile validity. On this point, Butcher, Cabiya, Lucio, and Garrido (2007) have found that both F scale and the L scale scores tend to be slightly higher in nonclinical samples of Hispanic Americans with low levels of acculturation than those who are highly acculturated. Butcher et al. recommended increasing cut scores on the MMPI-2 allowing a slightly higher elevation (5 T-score points) on feigning scales before considering a profile invalid. Currently, little research exists on the interpretation of defensiveness scales. This scarcity of studies could be due to the previously discussed overlap in the constructs of defensiveness and socially desirable responding addressed by Cashel et al. (1995).

Overall, very little research has examined clinical differences between Hispanic Americans and European Americans on MMPI-2 scales. With most of the research having been conducted on undergraduate students with presumably low levels of psychopathology, Greene (2000) cautions against making general statements about the cultural response styles of Hispanic American patients on the MMPI-2. He concluded that doing so is premature for this clinical population and further research is necessary before applying research findings to clinical assessment results.
Patterns of PAI elevations for Hispanic Americans are more problematic for validity indicators than clinical scales. Regarding the latter, studies have examined the clinical utility of the Spanish-language version of the PAI, and found it to be moderately effective for identifying major depression and schizophrenia (Fantoni-Salvador & Rogers, 1997). Research has also found good test-retest reliability for the Spanish PAI (Fernandez et al., 2008; Rogers, Flores, Ustad, & Sewell, 1995).

Research has raised serious concerns regarding the usefulness of Spanish PAI validity scales (Rogers et al., 1995). In a clinical sample, Romain (2000) found that more than 40% of the PAI protocols from Hispanic Americans were considered “invalid” based on the standard cut scores outlined in the PAI manual (Morey, 1991, 2007), twice as many as European American protocols. On average, Hispanic Americans had substantially higher PIM scores as compared to European Americans (Cohen’s $d = .60$). Hopwood, Flato, Ambwani, Garland, and Morey (2009) also found increased socially desirable response styles in Hispanic American non-clinical populations. Hispanic American undergraduates consistently attained higher scores than European Americans on three scales of socially desirable responding. Effect sizes were generally small: Defensiveness index (DEF; $d = .28$), Cashel Discriminant Function (CDF; $d = .37$), and Positive Impression Management (PIM; $d = .13$). Neither of these studies included measures of acculturation, so it is impossible to determine which cultural characteristics, if any, contributed to higher PAI scores for Hispanic Americans.

Despite cultural differences, average levels of defensiveness in the Romain (2000) study actually appear to demonstrate relatively little defensiveness with mean
PIM scores of 45.32 for Hispanic Americans and 38.06 for European Americans after invalid protocols were excluded from analysis. Unfortunately, DEF and CDF were not analyzed. Nonetheless, it is generally misleading to only consider mean values for any scale because within-culture differences for minority groups can be obscured when acculturation is not assessed (Anastasi, 1988; Berry, 1989). Additionally, the exclusion of invalid profiles due to high PIM scores, obfuscates the meaning of Romain’s results. There is the distinct possibility that culturally-specific response styles led to PIM elevations and Romain’s analyses were limited to PIM means for Hispanic American individuals with similar levels of acculturation to European Americans. Previously discussed elevations in scales designed to evaluate defensiveness and socially desirable responding raise the strong possibility that Hispanic Americans are increasingly reticent to disclose information related to treatment issues compared to individuals from other cultural groups.

Furthering the hypothesis of increased defensiveness among Hispanic American populations, Fernandez et al. (2008) found evidence of possible defensiveness within their sample, as individuals responding honestly exhibited a greater tendency to underreport symptoms on the Spanish version, particularly on CDF ($M_{\text{Spanish}} = 61.48$, $SD = 9.96$; $M_{\text{English}} = 56.40$, $SD = 6.33$). Although Fernandez et al. used a within-subjects design with bilingual individuals to compare English and Spanish PAI versions, Hopwood et al. (2009) also found CDF had the largest effect size in their study comparing European Americans to Hispanic Americans. Unfortunately, Romain (2000) did not compare validity scale scores for Hispanic Americans and European Americans.
in her study, so their results cannot be discussed in relation to other published studies on the Spanish PAI.

Few studies follow the ITC guidelines (Hambleton, 2001; Weiss & Rosenberg, 2012) and evaluate linguistic equivalence of English and Spanish versions of the PAI by administering both versions to bilingual participants. In a study using bilingual university students and non-patient community members, Fernandez et al. (2008) noted that validity scales on the English and Spanish PAI versions showed relatively equivalent levels of performance when differentiating between honest responders and individuals asked to feign or respond defensively. For the PIM scale, Fernandez et al. found moderately high English to Spanish correlations of .77 for honest responders and .78 for those in the under-reporting condition. These correlations are in stark contrast to the PIM correlation of .21 found by Rogers, Flores, Ustad, and Sewell (1995) in a population of Hispanic American patients.

In addressing the disparity between these two studies, Fernandez et al. noted that marked differences in linguistic equivalence may contribute to differences in the samples of the two Spanish PAI studies. Specifically, Fernandez et al. (2008) utilized a non-clinical, better educated sample than the Rogers et al. (1995) clinical outpatient sample. Furthermore, Rogers et al. did not screen participants for reading ability, nor did they exclude profiles demonstrating inconsistent responding. Neither study examined level of acculturation, so it is not possible to determine whether that also played a role in the disparity between the two studies. A final factor could be that there are qualities specific to the PIM scale that limit its effectiveness and stability among certain samples of Hispanic American individuals. Specifically, Rogers et al. (1995)
found a modest correlation of .21 for the PIM scale, but much higher correlations for the remaining validity scales (i.e., INC, INF, and NIM), which ranged from .58 to .83. PIM was also identified as having the smallest effect size \((d = .13)\) when differentiating between Hispanic American and European American students (Hopwood et al., 2009).

**Spanish SIRS-2**

A simulation study using the Spanish SIRS-2 identified a cultural response pattern that may be significant for the detection of malingering. Correa and Rogers (2010) compared Hispanic American outpatients with Traditional levels of acculturation to those at other levels. Interestingly, Traditional individuals exhibited a slightly higher than average effect size for amplified detection strategies than unlikely detection strategies \((M_d = 2.13\ vs\ M_d = 2.01)\). In contrast, Hispanic American individuals in the English SIRS-2 validation sample (Rogers, Sewell, & Gillard, 2010) demonstrated larger effect sizes for unlikely detection strategies and evidenced higher effect sizes on these strategies than European American individuals. Such differences between the English-speaking Hispanic American sample and the Spanish-speaking sample were expected because the English-speaking sample likely differs significantly in level of acculturation from the predominantly Traditional Spanish-speaking sample. Although Lucio et al. (2002) did not assess for acculturation in their sample, they hypothesized cultural differences in defensiveness as a primary cause for differences in responding on the MMPI-A between Hispanic American adolescents in the United States and adolescents in Mexico.
The differences in the detection strategies discussed above are small. However, if future research also demonstrates this pattern, it could suggest that strategies using the report of plausible symptoms to an exaggerated degree may be slightly more effective in distinguishing Traditionally-oriented feigners from honest responders due to cultural factors. For example, these findings could indicate that Traditional Hispanic American individuals have more difficulty identifying symptoms that European American individuals consider to be uncommon or unlikely, making them less prone to endorse these items when attempting to malinger. Alternatively, smaller effect sizes for the unlikely detection strategies might reflect defensiveness—even in the feigning condition—and a reticence to endorse symptoms of extreme pathology. The fact that 30% of participants in the honest condition attained scores that indicate defensiveness on the SIRS-2 DS subscale further corroborates the possibility of a culturally specific response style relating to defensiveness.

Linguistic and Cultural Considerations when Using the Spanish PAI

The effects of language are vitally important to consider when determining the accuracy of the assessment process. First of all, the psychometric properties of standardized assessment measures are likely to change when administered to individuals who are culturally different from the normative sample (Marin & Marin, 1991). Furthermore, multilingual individuals that are not tested in their preferred language can suffer a detachment effect (Bamford, 1991) and fail to adequately connect with assessment questions or be able to fully express their emotional and psychological issues. The detachment effect can result in poor communication about symptoms and
less self-disclosure (Dana, 1995), potentially magnifying the appearance of defensive response styles. This detachment effect is often remedied when individuals are tested in their preferred language. For example, Guttfreund (1990) shows that bilingual Hispanic American patients who prefer to speak Spanish are more able to effectively express their emotions when tested in this language rather than English. For the Spanish PAI, clinicians must take into account a client’s language preference prior to beginning the assessment process. Depending on the validation of the Spanish translation, the Spanish version may be the most appropriate when a strong preference is expressed for Spanish or the individual’s English language abilities are limited.

Although the PAI test manuals (Morey, 1991, 2003) do not describe the translation process for the Spanish version, it is available on their website (see http://www3.parinc.com/dynspage.aspx?PageCategory=Permissions&id=2). Its publishing company, Psychological Assessment Resources, has standardized their translation process, following the recommendations of most researchers and requiring an independent back-translation with review and approval by the test's author (Correa & Rogers, 2010; Marin & Marin, 1991).

Validation studies indicate very good test-retest reliability for the clinical scales of the Spanish PAI with monolingual patients (\(M_r = .78\)) and moderately good test-retest reliability between English and Spanish administrations for bilingual patients (\(M_r = .71\); Rogers et al., 1995). Good convergent validity has been found for the PAI with the Spanish version of the Diagnostic Interview Schedule (DIS; Robins, Helzer, Croughan, & Ratcliff, 1981) regarding symptoms of major depression, schizophrenia, alcohol
dependence, and anxiety disorders (Fantoni-Salvador & Rogers, 1997). These results indicate good diagnostic accuracy for the Spanish language version of the PAI.

Finally, psychological research with Hispanic American patients must take into account important cultural differences among individuals with different countries of origin (Puente, 1990). Spanish PAI results have been compared for Puerto Ricans, Mexican Americans, and Latin Americans, finding no significant between-group differences in PAI response patterns (Fantoni-Salvador & Rogers, 1997). Examination of this issue helps minimize the concern of imposed etics because it is not assumed that all Hispanic cultures will have similar response patterns (Berry, 1988).

Purpose of the Current Study

The current study evaluated whether the Spanish PAI effectively distinguishes between Spanish-speaking outpatient groups randomly assigned to honest, feigning, and defensive conditions. Additionally, the study explored the role of acculturation on response styles among Spanish-speaking Hispanic American clinical populations and investigated the constructs of malingering and defensiveness as they apply to this clinical population. Lastly, the study tested the Bipolarity Hypothesis and investigated any potential effects of culturally specific response styles.

Research Questions and Hypotheses

1. Do the validity indicators of the Spanish PAI effectively differentiate honest responding outpatients in the standard (honest) condition from outpatients in the feigning and defensive conditions?

Consistent with past research (Fernandez et al., 2008; Morey, 2007), the first
research question tested whether higher elevations will be obtained on the Spanish PAI validity indicators for outpatients in the feigning and defensive conditions than those in the honest condition.

- Hypothesis 1: Outpatients in the feigning condition will achieve higher scores on the NIM, MAL, RDF, and NDS indicators of the Spanish PAI than outpatients in the honest condition.

- Hypothesis 2: Outpatients in the defensive condition will achieve higher scores on PIM, DEF, and CDF than outpatients in the honest condition.

2. How accurate are cut scores when applied to the Spanish PAI for classifying honest, feigning, and defensive conditions in a Spanish-speaking outpatient sample?

Current feigning research (Fernandez et al., 2008) on the Spanish PAI indicates NIM and PIM demonstrate high levels of accuracy among validity indices for the identification of simulators in a community sample. However, the generalizability of these results is limited to highly educated, non-clinical Hispanic Americans, and does not necessarily apply to monolingual patients. This research question sought to examine the utility of existing cut scores within a primarily monolingual Spanish-speaking clinical sample.

3. Do different levels of acculturation predict elevations on feigning and defensiveness indicators on the Spanish PAI?

This research question explored whether different levels of identification with American culture, based on scores from the Acculturation Rating Scale for Mexican Americans - 2nd edition (ARSMA-II; Cuellar, Arnold, & Maldonado, 1995), predict scores on NIM, MAL, RDF, NDS, PIM, DEF, and CDF on the Spanish PAI. Of particular interest was each outpatient’s linear Acculturation score, calculated using the ARSMA-II Anglo Orientation Subscale (AOS) and Mexican Orientation Subscale (MOS).
Acculturation scores place individuals on a continuum from Very Mexican-oriented to Very Anglo-oriented.

Hypothesis 3: Low acculturation scores will predict high scores on the PAI DEF.

4. Scores on PAI indicators of feigning and defensiveness will be inversely correlated.

According to the bipolarity hypothesis (Greene, 1997), scores on feigning scales and defensiveness scales should show an inverse relationship. This research question investigated whether scores on PAI NIM, MAL, RDF, and NDS are negatively correlated with scores on PIM, DEF, and CDF. This research question was analyzed by determining the strength of bivariate Pearson product-moment correlations between scale scores on feigning indicators and scale scores on defensiveness indicators for participants in the feigning and defensive conditions.

Hypothesis 4: Outpatients in the feigning and defensiveness conditions will have larger negative correlations between their respective validity indicators than those in the control condition.

Supplementary Question

Outpatients in the honest condition with different primary diagnostic categories (anxiety disorders, depressive disorders, and psychotic disorders) will have significantly different elevations on the validity scales of the Spanish PAI.

This supplementary question explored whether outpatients with different diagnostic categories exhibited different elevations on the validity scales of the Spanish PAI. Based on past research (Correa, 2010), three main symptom constellations can be analyzed from the sample: depression, anxiety, and psychotic disorders.
CHAPTER 2

METHODS

Study Design

The current study used a between-subjects simulation design with two experimental conditions (i.e., feigning and defensive) and one control condition (i.e., honest). Simulation designs allow researchers to test the utility of specific detection strategies for response style measures and scales. This design is commonly used in response style research because of its excellent internal validity (e.g., random assignment to groups). Because motivation for external gain is a crucial factor in the determination of malingering (APA, 2000), simulation studies typically offer participants external (e.g., monetary reward), or internal (e.g., the satisfaction of being told they “fooled the examiner” or “beat the test”) incentives for giving a convincing portrayal of a particular response style (Rogers, 2008). Accordingly, the current study utilized experimental scenarios, incentives, and asked participants to adopt a specific response style (Hawes & Boccaccini, 2009; Rogers, 1990; Rogers, 2008; Rogers & Gillard, 2010).

An additional component in simulation designs is the implementation of manipulation checks, which are essential in order to determine whether experimental instructions were adequately followed. More specifically, manipulation checks are used to ensure that participants understood the instructions, followed them, and maintained acceptable motivation throughout the study (Rogers & Gillard, 2010). Since motivation for response styles must be established, simulation designs cannot be considered effective for participants who do not sufficiently adopt the instructions for their assigned condition (Rogers & Gillard, 2010).
A final important consideration with simulation designs is whether relevant samples are utilized for the appropriate clinical comparisons. Towards this objective, participants in the current study were outpatients from Centro de Mi Salud, a treatment center designed specifically for Hispanic American patients, who need for mental health services provided in Spanish. These outpatients have direct knowledge of clinical services and a personal understanding of mental disorders that may assist them in how to portray or deny symptoms.

Participants

The initial sample was composed of 94 Spanish-speaking Hispanic outpatients, aged 18 years and older, that were recruited from Centro de Mi Salud, an outpatient mental health center in Dallas, Texas. Centro de Mi Salud specializes in providing low-cost mental health services to people of low socioeconomic status whose primary language is Spanish. Common diagnoses among patients at Centro de Mi Salud include mood, anxiety, and psychotic disorders.

To maintain the representativeness of the sample, inclusion criteria for the study were broad and inclusive whereas the exclusion criteria were minimal. Inclusion criteria were (a) adulthood (i.e., at least 18 years of age), (b) Spanish as their primary language, and (c) at least a fourth grade reading level as determined by the Reading Level Indicator (RLI; Williams, 1997). The only exclusion criterion was the presence of severe psychotic symptoms that impair the patients’ ability to understand and respond relevantly to the measures. In past research at the same setting (Correa, 2010), this exclusion criterion did not remove any participants.
Materials

Spanish Personality Assessment Inventory (PAI; Morey, 1991)

The Spanish PAI is a 344-item self-report designed to assess personality traits and symptoms of psychopathology. The measure contains 11 clinical scales, 5 treatment scales, and 2 interpersonal scales. In addition, the Spanish PAI contains 4 standard validity scales for measuring response style and profile validity (Morey, 1991, 2007).

According to Fernandez et al. (2008), the Spanish PAI clinical scales showed a moderate to good correspondence between Spanish and English versions ($M_r = .72$) and good test-retest reliability between Spanish language administrations ($M_r = .79$). Rogers and Flores (1995) also found the Spanish PAI demonstrated moderate correspondence between both language versions ($M_r = .68$). Additionally Rogers and Flores (1995) demonstrated generally adequate alpha coefficients for Spanish PAI clinical scales ($M = .68$; range from .40 to .82) and treatment and interpersonal scales ($M = .62$; range from .40 to .82). However, two clinical (i.e., ANT, and ALC) and two treatment scales (i.e., SUI and STR) lack good internal consistency (alphas <.60) among Spanish-speaking Hispanic American individuals.

English to Spanish correspondence on the PAI validity scales was moderately good reliability between English and Spanish administrations for bilingual patients ($M_r = .59$; Rogers et al., 1995). Test validity remains uninvestigated for individuals whose primary language is Spanish and who may have lower levels of acculturation. Researchers (Fernandez et al., 2008; Rogers et al., 1995) caution practitioners about using Spanish PAI validity scales without clear empirical support.
The Acculturation Rating Scales for Mexican Americans—2nd edition (ARSMA-II; Cuellar, Arnold, & Maldonado, 1995)

The ARSMA-II is among the most widely used and well researched acculturation scales (Gamst et al., 2002). It contains two subscales with good internal consistency: the Anglo Orientation Subscale (AOS; Cronbach’s alpha = .86) and the Mexican Orientation Subscale (MOS; Cronbach’s alpha = .88), both of which are combined to produce an overall rating describing a person’s degree of acculturation. One important advantage of the ARSMA-II is that its Spanish language version has been researched and validated for use with Spanish-speaking populations. This validation distinguishes it from other acculturation measures, whose psychometric properties have yet to be determined for Spanish translations (Malcarne, Chavira, Fernandez, & Liu, 2006).

Reading Level Indicator (RLI; Williams, 2000)

The RLI is a 40-item multiple choice screening test that assesses reading level. The Spanish language version of the RLI has demonstrated excellent reliability and internal consistency (alpha = .93) among a sample of bilingual college students and non-patient community members (Fernandez et al., 2008). According to the RLI manual, evidence for content validity was obtained by utilizing expert knowledge in the creation of items that test essential parts of reading ability. Evidence of construct validity stems from the rigorous selection criteria for test items and expert feedback regarding construct validity (Williams, 2000).

Demographics Questionnaire

This brief questionnaire asked outpatients to report their age, occupation,
gender, and ethnicity/race. It is included in Appendix A.

Procedure

The study received ethical approval from the Institutional Review Board (IRB) at the University of North Texas and administrative approval from Centro de Mi Salud. All participants were provided informed consent in Spanish for their involvement in the study. Potential participants were provided with written consent forms, which were also read aloud by the researcher. By adopting this procedure, issues of limited literacy were addressed without any perceived stigmatization for participants with low reading levels.

Informed consent and instructions for all parts of the study were explained to each participant individually in an office or conference room, depending on available space in the clinic. Participants were then allowed to choose whether to complete the self-reports in the nearby clinic waiting area or be seated in a chair directly outside the room occupied by the researcher. They were instructed to return to the researcher’s room after the completion of each questionnaire so they could receive instructions for the next part of the study. Participants were also encouraged to go to the researcher’s room to ask any questions they might have about their task. The researcher also checked on each participant at approximately 15-minute intervals to ensure they were adequately engaged in the task and determine if they had any questions or concerns.

Phase I

Following the written informed consent in Spanish, all participants were
evaluated by the researcher, a bilingual doctoral student. Each participant began by completing the demographics questionnaire and the RLI. Subsequently, each participant’s level of acculturation was assessed via the ARSMA-II, a self-report measure of their activities, and cultural preferences. The researcher scored each participant’s RLI while participants completed the ARSMA-II. Because the Spanish PAI requires at least a 4-grade reading level (Fernandez et al., 2008), it was only administered to participants whose RLI scores indicated reading proficiency at or above the fourth-grade level. Participants with reading abilities lower than a fourth grade level were thanked and excused from the study. Their initial data were excluded from further analysis.

After Phase I was completed, the researcher introduced participants to their Phase II conditions for the Spanish PAI, either the honest, feigning, or defensive condition. Prior to data collection, the three conditions and their instructions were shuffled and sealed into identical white envelopes. Envelopes were then placed into each testing packet in a quasi-random fashion. Neither the investigator nor the participant knew the experimental condition until the envelope was opened just prior to explaining the instructions. After the instructions were explained, participants were asked to paraphrase instructions to ensure comprehension; they also had an opportunity to ask questions before beginning Phase II. If they were unable to comprehend experimental instructions after asking questions and receiving additional explanation from the examiner, participants were excused from the study and their data were omitted from any subsequent analysis. It should be noted that no participants were excluded due to inability to comprehend the instructions for their condition.
Phase II

Participants were asked to complete the Spanish PAI according to their experimental instructions. For participants in the feigning and defensive conditions, this involvement required them to modify their answers based on their experimental scenario and accompanying instructions.

Scenarios

Simulation designs require that the feigning and defensive conditions be relevant to participants, engaging, and easily understood (Rogers & Cruise, 1998). For this reason, participants were presented with a scenario with which they are likely to have experience. Because all participants were established patients at Centro de Mi Salud, individuals in the feigning condition were asked to simulate persons who are intentionally fabricating or exaggerating symptoms to gain benefits and entry into a specific program at the treatment center (for full instructions, see Appendix B). The multiple benefits mentioned in the scenario were designed to be appealing to patients at this particular treatment center (e.g., free transportation, free treatment for self and family members, and preference in the scheduling of appointment times).

For the second scenario, participants in the defensive condition were asked to simulate people who are intentionally minimizing symptoms and attempting to present themselves as well-adjusted as possible in order to obtain the same benefits as above and gain entry into the hypothetical treatment program (see Appendix B for full instructions). Both experimental scenarios were kept as similar as possible, in order to
maintain consistency in the services between conditions. The simulated benefits mentioned in this scenario were the same as those described in the feigning scenario.

Participants in the feigning and defensive conditions were cautioned to be convincing in their presentations, and challenged to “fool the examiner” into believing they were responding truthfully in their portrayal of feigned or minimized symptomatology (Correa & Rogers, 2010). Such warnings are consistent with experimental instructions in past feigning research with Hispanic American patients. See Appendix B for full scenarios and all experimental instructions.

For the honest (control) condition, participants were asked to be truthful and forthcoming about their current symptoms. They were not presented with a scenario, because this could have potentially affected the genuineness of their responses. Instead, their instructions stressed the importance of this research in creating a valid test that would be of optimum use in helping Hispanic American patients undergoing psychological evaluations (see Appendix B for complete instructions). Additionally, these instructions stressed the importance of participants’ role in helping the Hispanic American community, by assisting in this research.

Manipulation Check

After completing all measures, the researcher conducted a manipulation check with each participant (see Appendix C). At this time, participants were asked to recall the experimental instructions in their own words as the researcher recorded their responses. Participants were also asked to rate how much effort they put forth in following their instructions. Participants were excluded from data analysis because of
limited adherence to the experimental condition if they: (a) could not remember their experimental instructions, (b) reported not following instructions, or (c) reported they “did not try very hard” to follow instructions. The specific questions posed to participants during the manipulation check can be found in Appendix C. After the manipulation check, all participants were debriefed and informed about the general goals of the study.

Procedure for the Exclusion of Invalid Profiles

As noted, previous research by Romain (2000) found that nearly 40% of Spanish PAI profiles were considered invalid based on the suggested validity scale cut scores in the PAI manual. Romain’s study excluded invalid profiles from analysis. However, the current study examines the effectiveness of different cut scores suggested across the literature for both English and Spanish versions of the PAI (Fernandez et al., 2008; Hawes & Boccaccini, 2009; Morey, 2007). Therefore, no PAI profiles were excluded from preliminary data analysis in the current study due to their scores on indicators of feigning or defensiveness.

To date, Spanish PAI studies have not examined the effects of ICN and INF scale scores on profile validity; nor have appropriate cut scores been suggested (Fernandez, Boccaccini, & Noland, 2007; Romain, 2000). The ICN and INF scales are designed to measure appropriate attention to item content, and high scores are indicative of possible carelessness, confusion, reading difficulties, or random responding to the PAI. Therefore, it is imperative that these scales be studied in a sample of Spanish-speaking patients. Currently, no established guidelines are
published for interpreting these scales with populations other than the English-speaking normative sample and clinical samples of the PAI. No participants were excluded from the current study based on their INF scores, because its content may be interpreted differently when presented in Spanish to Hispanic patients. Using Morey’s (2007) general guideline, only participants with ICN scores lying 2 standard deviations above the sample mean were considered significantly elevated and excluded from analysis. Unlike INF, ICN utilizes pairs of items with opposite content. Therefore, an endorsement of incompatible content via these pairs is evidence of inconsistent responding.
CHAPTER 3

RESULTS

Refinement of the Sample

The initial sample consisted of 94 male and female Spanish-speaking Hispanic American outpatients who completed their participation in the study. Consistent with inclusion criteria, all clinic patients over the age of 18, with a tested reading level of 4th grade and higher on the Reading Level Indicator (RLI), were eligible for participation in the study. Three female patients and two male patients were excused from further participation after the administration of screening measures because they failed to achieve a fourth-grade reading level equivalence on the RLI. The only other exclusion criterion was the presence of psychotic symptoms which interfered significantly with the patient’s ability to comprehend the study, provide informed consent, and answer questions without experiencing distress. No participants were excluded due to the presence of severe psychotic symptoms.

Questions in the manipulation check led to further refinement of the sample, inasmuch as one additional female participant was excluded from further analysis because she reported not following feigning instructions during her experimental condition. Since simulation designs rely on a participant’s adherence to experimental instructions, it is crucial to only analyze data provided by individuals who reported (a) following instructions, and (b) putting adequate effort towards following instructions. No additional participants were excluded due to reported level of effort.

Previous studies (Romain, 2000) have excluded participants from analysis when they failed to yield valid Personality Assessment Inventory (PAI) profiles based on cut
scores established in the PAI manual (Morey, 2007). However, a major goal of the current study was to evaluate the effectiveness of PAI cut scores for feigning and defensiveness. Therefore, no participants were excluded on the basis of their feigning or defensiveness indicators. For the purposes of the current study, PAI profiles were only considered invalid if (a) participants omitted 18 or more items or (b) had inconsistent profiles. According to Morey (2007), the omission of 18 or more PAI questions indicates the examinee did not answer sufficient questions to yield an interpretable protocol. Three additional patients (3.3%) were excluded from subsequent analyses due to the number of test answers they omitted. Finally, 4 participants were excluded because their ICN scores were 2 standard deviations above the sample mean, indicating they responded inconsistently to PAI items.

Demographic Data

The final sample consisted of 25 (29.8%) male and 59 (70.2%) female outpatients ranging in age from 18 to 70 ($M = 37.65, SD = 10.28$) years. Not surprisingly, the majority of participating outpatients (78 or 92.9%) reported being born outside the United States. Their country of origin was predominantly Mexico (70 or 83.3%) with smaller representations from other countries: U.S. (6 or 7.1%), El Salvador (4 or 4.8%), Honduras (1 or 1.2%), Nicaragua (1 or 1.2%), Puerto Rico (1 or 1.2%), and Peru (1 or 1.2%).

The vast majority (81 or 96.4%) reported Spanish as their first language, which is clearly understandable because clinical services at Centro de Mi Salud are provided primarily in Spanish. Nearly half of the sample (38 or 45.2%) also reported speaking
“some” English although only one fifth of the sample (17 or 20.2%) described themselves as bilingual in Spanish and English. Of six participants born within the U.S., four were considered 2nd generation, and the other two were either 3rd or 5th generation Hispanic Americans.

Participating outpatients’ level of education ranged considerably, from elementary school to bachelor degree levels, with an average education being two years of high school ($M = 10.17$, $SD = 3.46$). The majority of participants (75.0%) received no education in the United States and attained an average level of education of 9.73 years in their country of origin. The remaining participants completed an average of 11.95 years in school and their education ranged from 2 to 14 years in the United States and 0 to 12 years in Latin American countries. The varied levels of education found in the current sample allow this study the unique opportunity to examine the potential effects of reading level and education on self-report scores. Previous Spanish PAI studies have either not evaluated level of education (Romain, 2000), or utilized university samples with high levels of education (Fernandez et al., 2007).

As summarized in Table 4, male and female outpatients had comparable backgrounds. Although not statistically significant because of limited power, males tended to be older ($d = 0.32$) and moved to the United States at an older age ($d = 0.39$).
Table 4

*A Comparison of Male and Female Hispanic American Outpatients on Demographic Variables*

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 25)</td>
<td>(n = 59)</td>
<td>(M)</td>
<td>(SD)</td>
<td>(M)</td>
<td>(SD)</td>
</tr>
<tr>
<td>Age</td>
<td>40.00</td>
<td>36.70</td>
<td>12.18</td>
<td>9.30</td>
<td>1.87</td>
<td>0.18</td>
</tr>
<tr>
<td>Age moved to U.S.(^a)</td>
<td>25.87</td>
<td>22.07</td>
<td>11.92</td>
<td>8.73</td>
<td>2.43</td>
<td>0.12</td>
</tr>
<tr>
<td>Years lived in U.S.</td>
<td>14.35</td>
<td>14.78</td>
<td>10.02</td>
<td>8.04</td>
<td>.04</td>
<td>0.84</td>
</tr>
<tr>
<td>Acculturation score</td>
<td>-1.58</td>
<td>-1.95</td>
<td>1.17</td>
<td>1.05</td>
<td>2.09</td>
<td>0.15</td>
</tr>
<tr>
<td>Reading level</td>
<td>9.38</td>
<td>9.90</td>
<td>3.36</td>
<td>2.90</td>
<td>.51</td>
<td>0.48</td>
</tr>
</tbody>
</table>

*Notes.* The Acculturation score is calculated using the ARSMA-II Anglo Orientation Subscale (AOS) and Mexican Orientation Subscale (MOS). Acculturation scores place individuals on a continuum from Very Mexican-oriented to Very Anglo-oriented. For males, \(n = 23\). For females, \(n = 55\). \(^a\)Six participants born in the United States are excluded from this analysis.

Overall, most patients (64.8%) moved to the United States as adults and had resided there for more than a decade. Their Spanish reading abilities tended to be much higher than the minimum grade level required by the study. However, these numbers were skewed by the inclusion of several participants with advanced educations.

Gender differences in defensiveness were explored in Table 5 for those in the honest condition. However, these findings were constrained by the limited power. Of the three PAI defensiveness indicators, only DEF evidenced a non-significant trend with males having nearly double the score of their female counterparts. While not statistically significant because of limited power, it still produced a moderate effect size.
Table 5

A Comparison of Male and Female Honest Responding Outpatients on PAI Validity Indicators

<table>
<thead>
<tr>
<th></th>
<th>Male (n = 11)</th>
<th>Female (n = 17)</th>
<th>F</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PAI Malingering Indicators</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIM</td>
<td>68.99</td>
<td>23.59</td>
<td>68.80</td>
<td>20.81</td>
<td>.00</td>
</tr>
<tr>
<td>MAL</td>
<td>57.19</td>
<td>14.57</td>
<td>53.76</td>
<td>11.32</td>
<td>.49</td>
</tr>
<tr>
<td>RDF</td>
<td>61.43</td>
<td>14.35</td>
<td>61.75</td>
<td>12.04</td>
<td>.00</td>
</tr>
<tr>
<td>NDS</td>
<td>11.09</td>
<td>9.32</td>
<td>11.65</td>
<td>7.80</td>
<td>.03</td>
</tr>
<tr>
<td><strong>PAI Defensiveness Indicators</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIM</td>
<td>47.96</td>
<td>15.62</td>
<td>45.66</td>
<td>14.76</td>
<td>.16</td>
</tr>
<tr>
<td>CDF</td>
<td>146.49</td>
<td>17.21</td>
<td>147.08</td>
<td>14.80</td>
<td>.01</td>
</tr>
<tr>
<td>DEF</td>
<td>3.45</td>
<td>2.42</td>
<td>1.94</td>
<td>1.75</td>
<td>3.70</td>
</tr>
<tr>
<td><strong>Other Validity Scales</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICN</td>
<td>63.16</td>
<td>12.49</td>
<td>66.32</td>
<td>11.34</td>
<td>.42</td>
</tr>
<tr>
<td>INF</td>
<td>59.81</td>
<td>8.84</td>
<td>59.11</td>
<td>13.41</td>
<td>.02</td>
</tr>
</tbody>
</table>

Note. For indicators, NIM = Negative Impression Management; MAL = Malingering Index; RDF = Rogers Discriminant Function; NDS = Negative Distortion Scale; PIM = Positive Impression Management; CDF = Cashel’s Discriminant Function; DEF = Defensiveness Index; ICN = Inconsistency Scale; INF = Infrequency Scale.

Effectiveness of the Spanish PAI Validity Indicators

PAI Validity Indicators

The discriminability of PAI validity indicators for specific response styles are critically important to their clinical usefulness. Hypotheses 1 and 2 predicted outpatients in the feigning condition would produce higher Spanish PAI scores on feigning indicators than those in the honest condition. Additionally, it is expected that individuals
in the defensive condition will produce higher scores on defensiveness indicators than honest responders.

Table 6

*Differences on the Spanish PAI Validity Indicators Between Honest and Feigned Presentations*

<table>
<thead>
<tr>
<th>PAI scales</th>
<th>Feigned ($n = 28$)</th>
<th>Honest ($n = 28$)</th>
<th>$F$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIM</td>
<td>97.44 26.10</td>
<td>68.87 21.51</td>
<td>19.98***</td>
<td>1.19</td>
</tr>
<tr>
<td>MAL</td>
<td>69.30 18.29</td>
<td>55.11 12.55</td>
<td>11.47***</td>
<td>0.90</td>
</tr>
<tr>
<td>RDF</td>
<td>70.95 13.37</td>
<td>61.61 12.75</td>
<td>6.13*</td>
<td>0.72</td>
</tr>
<tr>
<td>NDS</td>
<td>22.68 8.34</td>
<td>11.43 8.27</td>
<td>25.71***</td>
<td>1.35</td>
</tr>
<tr>
<td>INF</td>
<td>75.23 14.04</td>
<td>59.38 11.65</td>
<td>21.12***</td>
<td>1.23</td>
</tr>
</tbody>
</table>

For $F$ ratios, *$p < .05$, **$p < .01$, ***$p < .001$.

According to Rogers (2008) guidelines for malingering research, (a) moderate effect sizes are $d > 0.75$, (b) large effect sizes are $d > 1.25$, and (c) very large, $d > 1.50$). Spanish PAI validity indicators generally produced moderate to large effect sizes ($M d = 1.08$; range from 0.72 to 1.35). As seen in Table 6, PAI indicators utilizing Rare Symptoms strategies (NIM and NDS) demonstrated moderate to large effect sizes. In contrast, the Spurious Patterns strategies (MAL and RDF) which focus on patterns of response that are characteristic of malingering, but are very uncommon in clinical populations (MAL and RDF), appeared to be generally less effective with $ds < 1.00$.

The discriminability of validity scales was also explored for PAI measures of defensiveness and socially desirable responding. Specifically, the PIM, DEF, and CDF are designed to detect individuals, who are denying negative characteristics or
otherwise attempting to present themselves in an overly positive light. Spanish PAI validity indicators demonstrated moderate to very large effect sizes ($M d = 1.27$; range from 0.94 to 1.68). Notably, CDF produced the smallest effect size ($d = 0.94$) of all Spanish PAI validity indicators, including INF ($d = 0.94$). This finding is unexpected because, while the CDF uses 6 different scales to create a function score, it has been found to be more accurate in detecting defensiveness in the English version of the PAI than either the PIM or DEF scores alone (Cashel et al., 1995; Morey, 2007).

Table 7

* Differences on the Spanish PAI Validity Indicators Between Honest and Defensive Presentations *

<table>
<thead>
<tr>
<th>PAI scales</th>
<th>Defensive ($n = 28$)</th>
<th>Honest ($n = 28$)</th>
<th>$F$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>PIM</td>
<td>65.40</td>
<td>10.36</td>
<td>46.56</td>
<td>14.86</td>
</tr>
<tr>
<td>DEF$^a$</td>
<td>5.89</td>
<td>1.87</td>
<td>2.54</td>
<td>2.13</td>
</tr>
<tr>
<td>CDF$^a$</td>
<td>159.68</td>
<td>11.39</td>
<td>146.85</td>
<td>15.48</td>
</tr>
<tr>
<td>INF</td>
<td>75.78</td>
<td>20.37</td>
<td>59.38</td>
<td>11.65</td>
</tr>
</tbody>
</table>

Notes. For $F$ ratios, *$p < .05$, **$p < .01$, ***$p < .001$. $^a$ T score conversions could not be calculated for these indicators. Values are presented as raw scores.

significant differences in INF scores between groups suggest the possibility of idiosyncratic responding among Hispanic American patients both underreporting and overreporting symptoms on the Spanish PAI. Properties of the INF scale for the Spanish PAI and the possibility of a culturally-specific response style have not been researched, to date. A further investigation of INF items is shown in Table 8. Specifically, INF Item 40 shows a notable discrepancy between the honest and defensive conditions, with no honest responders endorsing the item. Item 320 also
attained a notably higher average score among participants in the malingering condition than for those in both the honest and defensive conditions.

Table 8

Mean Values for INF Item Endorsement by Hispanic American Outpatients on the Spanish PAI for Honest, Malingering, and Defensive Conditions

<table>
<thead>
<tr>
<th>INF Item Number</th>
<th>Summary of Item Content</th>
<th>Honest $M$</th>
<th>Malingering $M$</th>
<th>Defensive $M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>Favorite poet</td>
<td>0.00</td>
<td>0.72</td>
<td>1.04</td>
</tr>
<tr>
<td>80</td>
<td>Receiving unwanted ads in the mail</td>
<td>1.42</td>
<td>1.44</td>
<td>2.00</td>
</tr>
<tr>
<td>120</td>
<td>Favorite sport</td>
<td>0.27</td>
<td>0.84</td>
<td>1.07</td>
</tr>
<tr>
<td>160</td>
<td>Winning vs. losing</td>
<td>0.42</td>
<td>1.20</td>
<td>0.85</td>
</tr>
<tr>
<td>200</td>
<td>Favorite hobbies</td>
<td>0.12</td>
<td>0.92</td>
<td>1.04</td>
</tr>
<tr>
<td>240</td>
<td>Buying things that are overpriced</td>
<td>1.12</td>
<td>1.40</td>
<td>1.37</td>
</tr>
<tr>
<td>280</td>
<td>Looking forward to the dentist</td>
<td>1.15</td>
<td>0.72</td>
<td>1.48</td>
</tr>
<tr>
<td>320</td>
<td>How to spend free time</td>
<td>0.35</td>
<td>2.12</td>
<td>0.52</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.61</td>
<td>1.17</td>
<td>1.17</td>
</tr>
</tbody>
</table>

Utility of Spanish PAI Scales

The overarching goal of Research Question 2 was to investigate the accuracy of PAI cut scores for distinguishing the two simulation conditions from outpatients in the honest condition. The effectiveness of cut scores suggested in English PAI studies were evaluated using those included in the PAI manual (Morey, 2007), and in a recent PAI meta-analysis by Hawes and Boccaccini (2009). Regarding the Spanish PAI, only
one study has suggested optimal cut scores to date (Fernandez et al., 2008). Using a non-clinical sample of bilingual Hispanic American individuals, Fernandez et al.'s values are designed to maximize the Overall Correct Classification (OCC), a general measure of the overall accuracy of the test. In contrast to Fernandez et al. (2008), the relative effectiveness of each suggested cut score was assessed for this sample, error rates were calculated, and additional cut score values were tested.

Although sensitivity and specificity are commonly used, a brief review of other utility estimates is beneficial. Positive predictive power (PPP) is the proportion of those classified as feigning, who are correctly identified, whereas the negative predictive power (NPP) is the proportion of those classified as not feigning, who are correctly identified. The base rate refers to the frequency with which something (e.g., malingering) typically occurs. Both PPP and NPP can also be calculated for different base rates. In the current study, outpatients were randomly assigned to experimental conditions of nearly equal group size. Therefore, the base rate of malingering for the current study is artificially high at approximately 50%. In clinical and forensic populations, base rates vary widely, but are much lower than 50% (Rogers, 2008). Rogers et al. (1998) found base rates for malingering ranged from 10 – 30% (SD = 14.4). Therefore, the current study sought to examine base rates near the midpoint of these percentages (i.e., 15% and 25%). This percentage also represents the midpoint for PAI research by Rogers, Gillard, Wooley, and Kelsey (2012), who examined base rates of 15% and 25% to evaluate the effectiveness of cut scores for feigned mental disorders.
As Table 9 illustrates, utility estimates were employed to identify likely feigners on the Spanish PAI. They were tested using the criteria set forth in the PAI manual (Morey, 2007) and adjusted to minimize false positives (e.g., NPP ≥ .95).

Table 9

*Utility of PAI Feigning Indicators for Differentiating between Likely Genuine and Likely Feigning Responders*

<table>
<thead>
<tr>
<th>PAI Indicator</th>
<th>Sens</th>
<th>Spec</th>
<th>OCC</th>
<th>PPP (BR = 15%)</th>
<th>NPP (BR = 15%)</th>
<th>PPP (BR = 25%)</th>
<th>NPP (BR = 25%)</th>
<th>PPP (BR = 50%)</th>
<th>NPP (BR = 50%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Likely Genuine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIM ≤ 70T</td>
<td>.82</td>
<td>.61</td>
<td>.71</td>
<td>.27</td>
<td>.95</td>
<td>.41</td>
<td>.91</td>
<td>.68</td>
<td>.77</td>
</tr>
<tr>
<td>NIM ≤ 77T</td>
<td>.82</td>
<td>.68</td>
<td>.75</td>
<td>.31</td>
<td>.96</td>
<td>.46</td>
<td>.92</td>
<td>.72</td>
<td>.79</td>
</tr>
<tr>
<td><strong>Likely Feigning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIM ≥ 81T&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.64</td>
<td>.79</td>
<td>.71</td>
<td>.35</td>
<td>.93</td>
<td>.50</td>
<td>.87</td>
<td>.75</td>
<td>.69</td>
</tr>
<tr>
<td>NIM ≥ 92T</td>
<td>.50</td>
<td>.82</td>
<td>.66</td>
<td>.33</td>
<td>.90</td>
<td>.48</td>
<td>.83</td>
<td>.74</td>
<td>.62</td>
</tr>
<tr>
<td>NIM ≥ 110T</td>
<td>.32</td>
<td>.93</td>
<td>.63</td>
<td>.45</td>
<td>.89</td>
<td>.60</td>
<td>.80</td>
<td>.82</td>
<td>.58</td>
</tr>
<tr>
<td>NIM ≥ 115T</td>
<td>.29</td>
<td>1.00</td>
<td>.64</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Likely Genuine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAL &lt; 1</td>
<td>.86</td>
<td>.36</td>
<td>.61</td>
<td>.19</td>
<td>.94</td>
<td>.31</td>
<td>.89</td>
<td>.57</td>
<td>.72</td>
</tr>
<tr>
<td><strong>Likely Feigning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAL ≥ 3&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.25</td>
<td>.96</td>
<td>.61</td>
<td>.52</td>
<td>.88</td>
<td>.68</td>
<td>.79</td>
<td>.86</td>
<td>.56</td>
</tr>
<tr>
<td>MAL ≥ 4</td>
<td>.25</td>
<td>.96</td>
<td>.61</td>
<td>.52</td>
<td>.88</td>
<td>.68</td>
<td>.79</td>
<td>.86</td>
<td>.56</td>
</tr>
<tr>
<td><strong>Likely Genuine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDF ≤ 59T</td>
<td>.75</td>
<td>.57</td>
<td>.66</td>
<td>.24</td>
<td>.93</td>
<td>.37</td>
<td>.87</td>
<td>.64</td>
<td>.70</td>
</tr>
<tr>
<td>RDF ≤ 60T&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.75</td>
<td>.57</td>
<td>.66</td>
<td>.24</td>
<td>.93</td>
<td>.37</td>
<td>.87</td>
<td>.64</td>
<td>.70</td>
</tr>
<tr>
<td>RDF ≤ 70T</td>
<td>.54</td>
<td>.71</td>
<td>.63</td>
<td>.25</td>
<td>.90</td>
<td>.38</td>
<td>.82</td>
<td>.65</td>
<td>.61</td>
</tr>
<tr>
<td><strong>Likely Feigning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDF ≥ 90T</td>
<td>.07</td>
<td>1.00</td>
<td>.54</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Likely Genuine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDS ≤ 11</td>
<td>.89</td>
<td>.57</td>
<td>.73</td>
<td>.27</td>
<td>.97</td>
<td>.41</td>
<td>.94</td>
<td>.67</td>
<td>.84</td>
</tr>
<tr>
<td>NDS ≤ 13</td>
<td>.86</td>
<td>.61</td>
<td>.73</td>
<td>.28</td>
<td>.96</td>
<td>.42</td>
<td>.93</td>
<td>.69</td>
<td>.81</td>
</tr>
<tr>
<td>NDS ≤ 18</td>
<td>.79</td>
<td>.75</td>
<td>.77</td>
<td>.36</td>
<td>.95</td>
<td>.52</td>
<td>.91</td>
<td>.76</td>
<td>.78</td>
</tr>
<tr>
<td><strong>Likely Feigning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDS ≥ 24</td>
<td>.54</td>
<td>.93</td>
<td>.73</td>
<td>.67</td>
<td>.91</td>
<td>.79</td>
<td>.84</td>
<td>.92</td>
<td>.64</td>
</tr>
<tr>
<td>NDS ≥ 25</td>
<td>.46</td>
<td>.96</td>
<td>.71</td>
<td>.67</td>
<td>.91</td>
<td>.79</td>
<td>.84</td>
<td>.93</td>
<td>.64</td>
</tr>
</tbody>
</table>

*Notes.* For cut scores, *T* = *T* score. For indicators, NIM = Negative Impression Scale; MAL = Malingering Index; RDF = Rogers Discriminant Function; NDS = Negative Distortion Scale. For utility estimates, BR = base rate; Sens = sensitivity; Spec = specificity; OCC = overall correct classification; PPP = positive predictive power; NPP = negative predictive power. <sup>a</sup> This superscript denotes Spanish PAI cut scores recommended by Fernandez and Boccaccini (2008) to optimize Overall Correct Classification (OCC).
As reported by Rogers et al. (2012), PAI cut scores can be utilized to rule-out feigning (i.e., a high likelihood that the PAI is not feigned) and rule-in feigning (i.e., a high likelihood that the PAI is feigned). For the purposes of this dissertation, the rule-out category will be referred to as “likely genuine,” and the rule-in category as “likely feigning.” For likely genuine cut scores, high levels of sensitivity and NPP are required. NDS ≤ 11 demonstrates an NPP approaching 1.00 and a sensitivity approaching 0.90, indicating likely genuine scores. For likely feigning, high levels of specificity and PPP are required. NIM ≥ 115T yields a perfect specificity and PPP of 1.00, which is consistent across base rates. In other words, all outpatients classified as feigning actually were instructed to malinger on the Spanish PAI. These scores indicate the NIM scale, which employs a Rare Symptoms detection strategy, produced the most effective rule-in cut scores at or above 115T. RDF > 90T, which is based on spurious patterns also performed very well with a base-rate of 15%.

No participants in this sample had MAL scores in the ≥ 5 range; so the cut score recommended by Morey (2007) could not be calculated. Notably, cut scores suggested by Hawes and Boccaccini (2009) to optimize the OCC also produced or tied for the highest OCC in this sample. However, these scores did not always prove optimal in the current study, because they did not minimize false positives (e.g., NPP ≥ .95).

Rogers and Bender (2012) discussed what they believe is a fundamental misassumption in the assessment of malingering: the laser accuracy of cut scores, where single point differences are used to classify response styles. Table 10 examines the accuracy of well-defined groups by removing “too-close-to-call cases” (i.e., an indeterminate group of ± 5T for feigning indicators and ±1 SEM (4T) for the NDS).
Table 10

Effectiveness of PAI Cut Scores for Feigning with the Exclusion of an Indeterminate Category

<table>
<thead>
<tr>
<th>Cut Scores</th>
<th>Likely Genuine</th>
<th>Likely Feigning</th>
<th>Likely Genuine</th>
<th>Likely Feigning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>Sens</td>
<td>Spec</td>
<td>OCC</td>
</tr>
<tr>
<td>NIM ≤ 70T (+5)</td>
<td>87.5</td>
<td>.88</td>
<td>.61</td>
<td>.76</td>
</tr>
<tr>
<td>NIM ≥ 77T (+5)</td>
<td>82.1</td>
<td>.78</td>
<td>.74</td>
<td>.76</td>
</tr>
<tr>
<td>NIM ≥ 81T (+5)</td>
<td>86.7</td>
<td>.78</td>
<td>.76</td>
<td>.77</td>
</tr>
<tr>
<td>NIM ≥ 92T (+5)</td>
<td>89.3</td>
<td>.58</td>
<td>.85</td>
<td>.72</td>
</tr>
<tr>
<td>NIM ≥ 110T (+5)</td>
<td>91.1</td>
<td>.32</td>
<td>1.00</td>
<td>.67</td>
</tr>
<tr>
<td>NIM ≥ 115T (+5)</td>
<td>92.9</td>
<td>.27</td>
<td>1.00</td>
<td>.63</td>
</tr>
<tr>
<td>RDF ≤ 60T (+5)</td>
<td>82.1</td>
<td>.77</td>
<td>.58</td>
<td>.67</td>
</tr>
<tr>
<td>RDF ≤ 70T (+5)</td>
<td>76.8</td>
<td>.48</td>
<td>.82</td>
<td>.65</td>
</tr>
<tr>
<td>RDF ≥ 90T (+5)</td>
<td>98.2</td>
<td>.04</td>
<td>1.00</td>
<td>.53</td>
</tr>
<tr>
<td>NDS ≤ 11 (+4)</td>
<td>75.0</td>
<td>.92</td>
<td>.53</td>
<td>.76</td>
</tr>
<tr>
<td>NDS ≤ 13 (+4)</td>
<td>76.8</td>
<td>.92</td>
<td>.63</td>
<td>.79</td>
</tr>
<tr>
<td>NDS ≤ 18 (+4)</td>
<td>75.0</td>
<td>.81</td>
<td>.90</td>
<td>.86</td>
</tr>
<tr>
<td>NDS ≥ 24 (+4)</td>
<td>67.9</td>
<td>.43</td>
<td>.96</td>
<td>.76</td>
</tr>
<tr>
<td>NDS ≥ 25 (+4)</td>
<td>71.4</td>
<td>.38</td>
<td>.96</td>
<td>.73</td>
</tr>
</tbody>
</table>

Notes. % = the percentage of sample retained for the classification when ± 5 or ± 1 SEM (i.e., ± 4) is removed; For utility estimates, BR = base rate; Sens = sensitivity; Spec = specificity; OCC = overall correct classification; PPP = positive predictive power; NPP = negative predictive power. a Superscripts denote Spanish PAI cut scores recommended by Fernandez and Boccaccini (2008) to optimize OCC.

Due to the restricted range, an indeterminate group could not be created for MAL cut scores. With the indeterminate group excluded, positive predictive power increased for nearly all feigning indicators at a base rate of 15%. In other words, following removal of “too-close-to-call” cases, the Spanish PAI was better able to accurately classify feigners. With the exclusion of the indeterminate group, negative predictive power also increased for NIM cut scores across base rates and for across NDS cut scores at base
rates of 15% and 50%. This increase in NPP indicates an increase in the PAI’s accuracy in classifying honest responders.

Well-defined NIM cut scores without too-close-to-call cases improved specificity to 1.00. This improvement was the most pronounced effect on optimal cut score upon removal of the indeterminate group. Specifically, Table 9 demonstrates NIM > 115T is the best indicator for individuals who are likely feigning (NPP = .89; PPP = 1.00; OCC = .63). With the indeterminate group removed (see Table 11), NIM > 110T becomes a slightly better indicator of likely feigners (NPP = .89, PPP = 1.00, OCC = .67).

Table 11

Errors in the Indeterminate Group for PAI Cut Scores on Malingering Indicators: False Alarms and False Misses at 50% Base Rate

<table>
<thead>
<tr>
<th>PAI Cut Scores</th>
<th>Indeterminate</th>
<th>False Positives</th>
<th>False Negatives</th>
<th>Overall Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likely Genuine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIM ≤ 70T (+ 5)</td>
<td>65 to 75</td>
<td>100.0</td>
<td>40.0</td>
<td>70.0</td>
</tr>
<tr>
<td>NIM ≤ 77T (+ 5)</td>
<td>72 to 82</td>
<td>37.0</td>
<td>0</td>
<td>18.8</td>
</tr>
<tr>
<td>Likely Feigning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIM ≥ 81T (+ 5)b</td>
<td>76 to 86</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NIM ≥ 92T (+ 5)</td>
<td>87 to 97</td>
<td>100.0</td>
<td>80.0</td>
<td>90.0</td>
</tr>
<tr>
<td>NIM ≥ 110T (+ 5)</td>
<td>105 to 115</td>
<td>67.0</td>
<td>100.0</td>
<td>83.4</td>
</tr>
<tr>
<td>NIM ≥ 115T (+ 5)</td>
<td>110 to 120</td>
<td>0</td>
<td>33.0</td>
<td>16.7</td>
</tr>
<tr>
<td>Likely Genuine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDF ≤ 60Ta (+ 5)</td>
<td>55 to 65</td>
<td>33.0</td>
<td>50.0</td>
<td>46.5</td>
</tr>
<tr>
<td>RDF ≤ 70T (+ 5)</td>
<td>65 to 75</td>
<td>44.0</td>
<td>50.0</td>
<td>45.9</td>
</tr>
<tr>
<td>Likely Feigning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDF ≥ 90T (+ 5)c</td>
<td>85 to 95</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Likely Genuine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDS ≤ 11 (+ 4)</td>
<td>7 to 15</td>
<td>71.0</td>
<td>18.0</td>
<td>35.4</td>
</tr>
<tr>
<td>NDS ≤ 13 (+ 4)</td>
<td>9 to 17</td>
<td>57.0</td>
<td>25.0</td>
<td>38.1</td>
</tr>
<tr>
<td>NDS ≤ 18 (+ 4)</td>
<td>14 to 22</td>
<td>42.0</td>
<td>33.0</td>
<td>55.0</td>
</tr>
<tr>
<td>Likely Feigning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDS ≥ 24 (+ 4)</td>
<td>20 to 28</td>
<td>10.0</td>
<td>62.0</td>
<td>40.0</td>
</tr>
<tr>
<td>NDS ≥ 25 (+ 4)</td>
<td>21 to 29</td>
<td>0</td>
<td>56.0</td>
<td>29.4</td>
</tr>
</tbody>
</table>

Notes. Overall Errors were calculated using unweighted averages. aDenotes Spanish PAI cut scores recommended by Fernandez and Boccaccini (2008). bAll scores in this range (NIM ≥ 81T [± 5]) were...
classified as Honest, so the “% of Errors” could not be calculated. There was only one participant whose scores fell within this range (RDF ≥ 90T [± 5]); therefore, the “% of Errors” could not be calculated.

Interestingly, the Spanish PAI cut scores which optimized the overall hit-rate in a sample of Spanish-speaking bilingual individuals (Fernandez et al., 2008) also optimized the overall classification rate in the current sample upon removal of individuals in the indeterminate range. This finding was not consistently the case prior to removal of the indeterminate group. As previously found, it also appears that feigning indicators utilizing rare symptoms detection strategies (items that are rarely endorsed by genuine patients) such as NIM and NDS produced the highest overall classification rates.

Scoring and interpretation practices for the PAI emphasize the utility of specific cut scores and encourage clinicians to employ the optimized cut scores most appropriate for their sample (Hawes & Boccaccini, 2009; Morey, 2007). However, Rogers et al. (2012) and Rogers and Bender (2012), caution practitioners about the high classification errors for indeterminate groups when utilizing single cut scores. Commonsensically, scores very close to the cut score are particularly vulnerable to classification errors (see Table 10).

Indeterminate cases were investigated to examine whether they should be considered as too-close-to-call (see Table 10). In general, errors in overall classification rate ranged from 16.7 – 90% for all feigning indicators. Misclassifications were particularly high for the NIM, with marked fluctuations across the cut scores evaluated. It should be noted that NIM ≥ 92T produced an overall error rate of 90%, but there was only one outpatient in the current sample whose score fell within this indeterminate
range. Therefore, the group size is likely insufficient for the purposes of calculating the effectiveness of this particular range.

Once again, scales based on rare symptoms strategies appear to be the most effective in correctly classifying malingers. This finding is especially true for cut scores above the previously identified rule-in marks. Specifically, using NDS ≥ 25 and NIM ≥ 115 no genuine individuals were misclassified, even within the indeterminate ranges. This result suggests NDS and NIM are, relatively, the best indicators to rely on for clinical practice.

PAI defensiveness indicators vary according to their levels of sensitivity, specificity, PPP and NPP and, consequently, vary in their effectiveness for accurately classifying response styles. For scores higher than the “likely defensive” cut scores, levels of defensiveness that affect the validity of a patient’s PAI profile should be strongly suspected. For example, PIM ≥ 72T demonstrates a positive predictive power of 1.0 for all base rates. All defensive outpatients were correctly classified as defensive were on the Spanish PAI. DEF and CDF only demonstrated clear “likely genuine” criteria for very low cut scores. Thus, guidelines for defensiveness on DEF and CDF are minimally acceptable for differentiating between likely genuine and likely defensive presentations. Due to the poor performance of CDF and DEF, PIM appears to be the most reliable scale for clinicians seeking to accurately identify defensive patients.

The overall classification rate for the cut scores suggested by Fernandez et al. (2008) did not generalize to the sample in the current study. Therefore, clinicians may wish to focus on the likely defensive cut scores identified in Table 12 when their clients share demographic characteristics close to those of the patients in the current sample.
This practice will minimize the likelihood that profiles from genuine patients will be mistakenly labeled as invalid due to scores on defensiveness indicators.

Table 12

Utility of PAI Defensiveness Indicators for Differentiating between Likely Genuine and Likely Defensive Responders

<table>
<thead>
<tr>
<th>PAI Indicator</th>
<th>Sens</th>
<th>Spec</th>
<th>OCC</th>
<th>PPP at different base rates</th>
<th>NPP at different base rates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BR = 15%</td>
<td>BR = 25%</td>
</tr>
<tr>
<td>Likely Genuine</td>
<td></td>
<td></td>
<td></td>
<td>PPP NPP</td>
<td>PPP NPP</td>
</tr>
<tr>
<td>PIM ≤ 57T</td>
<td>.79</td>
<td>.79</td>
<td>.79</td>
<td>.40 .96</td>
<td>.56 .92</td>
</tr>
<tr>
<td>Likely Defensive</td>
<td></td>
<td></td>
<td></td>
<td>BR = 50.0%</td>
<td>BR = 50.0%</td>
</tr>
<tr>
<td>PIM ≥ 61T</td>
<td>.68</td>
<td>.79</td>
<td>.73</td>
<td>.36 .93</td>
<td>.52 .88</td>
</tr>
<tr>
<td>PIM ≥ 64T</td>
<td>.54</td>
<td>.86</td>
<td>.70</td>
<td>.41 .91</td>
<td>.56 .85</td>
</tr>
<tr>
<td>PIM ≥ 70T</td>
<td>.46</td>
<td>.93</td>
<td>.70</td>
<td>.54 .91</td>
<td>.69 .84</td>
</tr>
<tr>
<td>PIM ≥ 72T</td>
<td>.29</td>
<td>1.00</td>
<td>.64</td>
<td>1.00 .89</td>
<td>1.00 .81</td>
</tr>
<tr>
<td>Likely Genuine</td>
<td></td>
<td></td>
<td></td>
<td>BR = 50.0%</td>
<td>BR = 50.0%</td>
</tr>
<tr>
<td>CDF &lt; 55T</td>
<td>1.00</td>
<td>.36</td>
<td>.68</td>
<td>.22 1.00</td>
<td>.34 1.00</td>
</tr>
<tr>
<td>Likely Defensive</td>
<td></td>
<td></td>
<td></td>
<td>BR = 15%</td>
<td>BR = 25%</td>
</tr>
<tr>
<td>CDF ≥ 70T</td>
<td>.21</td>
<td>.93</td>
<td>.57</td>
<td>.35 .87</td>
<td>.50 .78</td>
</tr>
<tr>
<td>Likely Genuine</td>
<td></td>
<td></td>
<td></td>
<td>BR = 50.0%</td>
<td>BR = 50.0%</td>
</tr>
<tr>
<td>DEF ≤ 2</td>
<td>1.00</td>
<td>.39</td>
<td>.70</td>
<td>.22 1.00</td>
<td>.35 1.00</td>
</tr>
<tr>
<td>Likely Defensive</td>
<td></td>
<td></td>
<td></td>
<td>BR = 50.0%</td>
<td>BR = 50.0%</td>
</tr>
<tr>
<td>DEF ≥ 4</td>
<td>.89</td>
<td>.71</td>
<td>.80</td>
<td>.35 .97</td>
<td>.51 .95</td>
</tr>
<tr>
<td>DEF ≥ 5</td>
<td>.79</td>
<td>.79</td>
<td>.79</td>
<td>.40 .96</td>
<td>.56 .92</td>
</tr>
<tr>
<td>DEF ≥ 6</td>
<td>.57</td>
<td>.89</td>
<td>.73</td>
<td>.48 .92</td>
<td>.63 .86</td>
</tr>
<tr>
<td>DEF ≥ 7</td>
<td>.36</td>
<td>.93</td>
<td>.64</td>
<td>.48 .89</td>
<td>.63 .81</td>
</tr>
</tbody>
</table>

Note. For cut scores, T = T score. For indicators, PIM = Positive Impression Management; CDF = Cashel’s Discriminant Function; DEF = Defensiveness Index. For utility estimates, BR = base rate; Sens = sensitivity; Spec = specificity; OCC = overall correct classification; PPP = positive predictive power; NPP = negative predictive power. aSuperscripts denote Spanish PAI cut scores recommended by Fernandez and Boccaccini (2008) to optimize Overall Correct Classification (OCC).

Due to the restricted range of CDF and DEF scores, an indeterminate group could not be created without removing a significant proportion of participants from each analysis. Therefore, only PIM cut scores could be evaluated.

For the PIM cut scores, the OCCs were notably much higher with the exclusion of indeterminate groups. Of particular note, sensitivity increased significantly from .79 to .94 for the “likely genuine” group. A concomitant increase in positive predictive power
for PIM “likely genuine,” demonstrates exclusion of the indeterminate range enables the Spanish PAI to better identify individuals responding defensively. With this exclusion, negative predictive power also increased for PIM cut scores across base rates. This increase in NPP indicates an increase in the PAI’s accuracy when classifying honest responders. The concurrent increase in specificity values also indicates PIM’s improved ability to correctly classify non-defensive individuals.

Table 13

*Effectiveness of PAI Cut Scores for Defensiveness Scales with the Exclusion of an Indeterminate Category*

<table>
<thead>
<tr>
<th>Cut Scores</th>
<th>%</th>
<th>Sens</th>
<th>Spec</th>
<th>OCC</th>
<th>PPP</th>
<th>NPP</th>
<th>PPP</th>
<th>NPP</th>
<th>PPP</th>
<th>NPP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likely Genuine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIM ≤ 57 (+ 5)</td>
<td>69.6</td>
<td>.94</td>
<td>.76</td>
<td>.85</td>
<td>.41</td>
<td>.99</td>
<td>.57</td>
<td>.97</td>
<td>.80</td>
<td>.93</td>
</tr>
<tr>
<td>Likely Defensive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIM ≥ 61 (+ 5)</td>
<td>75.0</td>
<td>.75</td>
<td>.91</td>
<td>.83</td>
<td>.60</td>
<td>.95</td>
<td>.74</td>
<td>.92</td>
<td>.89</td>
<td>.78</td>
</tr>
<tr>
<td>PIM ≥ 64 (+ 5)*</td>
<td>76.8</td>
<td>.68</td>
<td>.92</td>
<td>.81</td>
<td>.60</td>
<td>.94</td>
<td>.74</td>
<td>.90</td>
<td>.89</td>
<td>.74</td>
</tr>
<tr>
<td>PIM ≥ 70 (+ 5)</td>
<td>78.6</td>
<td>.35</td>
<td>1.00</td>
<td>.70</td>
<td>1.00</td>
<td>.90</td>
<td>1.00</td>
<td>.82</td>
<td>1.00</td>
<td>.61</td>
</tr>
<tr>
<td>PIM ≥ 72 (+ 5)</td>
<td>80.4</td>
<td>.32</td>
<td>1.00</td>
<td>.71</td>
<td>1.00</td>
<td>.89</td>
<td>1.00</td>
<td>.81</td>
<td>1.00</td>
<td>.60</td>
</tr>
</tbody>
</table>

**Notes.** % = the percentage of sample retained for the classification when + 5 or + 1 SEM is removed; For utility estimates, BR = base rate; Sens = sensitivity; Spec = specificity; OCC = overall correct classification; PPP = positive predictive power; NPP = negative predictive power. *Denotes Spanish PAI cut scores recommended by Fernandez and Boccaccini (2008).

Table 14 shows classification errors for individuals within the indeterminate ranges for PIM at various cut scores suggested in the literature. Errors in overall classification rate ranged from 31.1% to 63.9% for the identified PIM ranges. False positive rates were generally lower than false negative rates for each PIM cut score.
Notably, no honest responders were misclassified as yielding invalid protocols due to defensiveness at PIM ≥ 72 (False positive rate = 0%).

Table 14

*Errors in the Indeterminate Group for PAI Cut Scores: False Alarms and False Misses at 50% Base Rate*

<table>
<thead>
<tr>
<th>PAI Cut Scores</th>
<th>% of Errors</th>
<th>False Positives</th>
<th>False Negatives</th>
<th>Overall Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likely Genuine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIM ≤ 57 (+5)</td>
<td>52 to 62</td>
<td>17.0</td>
<td>45.0</td>
<td>31.1</td>
</tr>
<tr>
<td>Likely Defensive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIM ≥ 61 (+5)</td>
<td>56 to 66</td>
<td>50.0</td>
<td>67.0</td>
<td>58.4</td>
</tr>
<tr>
<td>PIM ≥ 64 (+5)*</td>
<td>59 to 69</td>
<td>50.0</td>
<td>78.0</td>
<td>63.9</td>
</tr>
<tr>
<td>PIM ≥ 70 (+5)</td>
<td>65 to 75</td>
<td>25.0</td>
<td>50.0</td>
<td>36.1</td>
</tr>
<tr>
<td>PIM ≥ 72 (+5)</td>
<td>67 to 77</td>
<td>0</td>
<td>78.0</td>
<td>38.9</td>
</tr>
</tbody>
</table>

*Note.* Overall errors were calculated using unweighted averages. *a* Denotes Spanish PAI cut score recommended by Fernandez and Boccaccini (2008).

**Internal Consistency of the Spanish PAI Validity Scales**

The internal consistency of Spanish PAI validity scales was investigated because they cannot be extrapolated from the original PAI. It is of vital importance to investigate internal consistency of Spanish PAI scales to help determine their scale homogeneity. As seen in Table 15, the alpha coefficients for each validity scale was acceptable (greater than .75), indicating that items within each scale measure the same general construct. Additionally, mean inter-item correlations are not so high as to indicate redundancy in test items. The current alpha values are generally comparable to the clinical standardization sample using the English PAI.
Table 15

Internal Consistencies and Standard Errors of Measurements (SEM) for the Spanish PAI Validity Scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>English Alpha&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Alpha</th>
<th>Mean Inter-Item r</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIM</td>
<td>.74</td>
<td>.76</td>
<td>.27</td>
<td>2.87</td>
</tr>
<tr>
<td>NDS</td>
<td>.74</td>
<td>.78</td>
<td>.22</td>
<td>3.84</td>
</tr>
<tr>
<td>PIM</td>
<td>.77</td>
<td>.76</td>
<td>.26</td>
<td>3.24</td>
</tr>
</tbody>
</table>

Notes. Because of their deliberate distortions, feigners are not expected to produce uniform results; therefore, SEMs are calculated using the alphas and SDs under the honest condition. <sup>a</sup>English alphas for NIM and PIM were reported by Morey (2007) for the clinical standardization sample. Alpha value for NDS was reported by Mogge et al. (2010).

Acculturation

The effects of acculturation on the Spanish PAI validity indicators was investigated in order to determine the generalizability of the Spanish PAI across primarily Spanish-speaking individuals who differ in their cultural identification (Anastasi, 1988; Okazaki & Sue, 1995; Wagner & Gartner, 1997). Research Question 3 sought to test the effects of acculturation on validity indicator scores.

ARSMA-II categories (e.g., Traditional, Marginal, Bicultural, and Acculturated) were not examined due to the cultural homogeneity of the sample, which was established by previous research at this site (Correa & Rogers 201). Instead, ARSMA-II scores were studied dimensionally and linear regression was used to investigate whether level of acculturation predicts scores on NIM, MAL, RDF, NDS, PIM, DEF, and CDF for honest participants on the Spanish PAI (see Table 16).
Table 16

**Acculturation as a Predictor for Scores on PAI Validity Indicators of Honest Responders**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE B</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIM</td>
<td>.87</td>
<td>3.59</td>
<td>.05</td>
</tr>
<tr>
<td>MAL</td>
<td>-4.56</td>
<td>1.90</td>
<td>-.43*</td>
</tr>
<tr>
<td>RDF</td>
<td>-2.25</td>
<td>2.22</td>
<td>-.21</td>
</tr>
<tr>
<td>NDS</td>
<td>-.43</td>
<td>1.38</td>
<td>-.06</td>
</tr>
<tr>
<td>PIM</td>
<td>-1.75</td>
<td>2.46</td>
<td>-.14</td>
</tr>
<tr>
<td>CDF</td>
<td>-2.14</td>
<td>2.55</td>
<td>-.16</td>
</tr>
<tr>
<td>DEF</td>
<td>-.27</td>
<td>.35</td>
<td>-.15</td>
</tr>
</tbody>
</table>

*p < 0.05

As seen in Table 16, the only significant relationship between validity indicators and ARSMA-II Acculturation Score proved to be a small negative association as evidenced by the MAL beta weight. That is, lower acculturation scores produced higher scores on MAL, indicating that MAL scores can be predicted based on acculturation level. The general lack of significant correlations suggests Spanish PAI validity indicators are relatively uninfluenced by acculturation. Although previous defensiveness research suggests culture affects defensiveness, these results indicate that varying levels of acculturation do not impact scores on the Spanish PAI.

**The Bipolarity Hypothesis**

According to the Bipolarity Hypothesis, malingering and defensiveness are considered to be two opposite endpoints on the same continuum. Therefore, scores on
these scales are expected show an inverse relationship (Greene, 1997). Research Question 4 posits that scores on the Spanish PAI NIM, MAL, and NDS are negatively correlated with scores on PIM, DEF, and CDF.

Table 17

*Pearson Correlation Matrix for Spanish PAI Validity Indicators among Hispanic American Outpatients in the Honest Condition*

<table>
<thead>
<tr>
<th></th>
<th>NIM</th>
<th>MAL</th>
<th>NDS</th>
<th>PIM</th>
<th>CDF</th>
<th>DEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIM</td>
<td></td>
<td>.58**</td>
<td>.81**</td>
<td>-.77**</td>
<td>.16</td>
<td>-.68**</td>
</tr>
<tr>
<td>MAL</td>
<td>-.56**</td>
<td></td>
<td>-.33</td>
<td>.29</td>
<td>-0.07</td>
<td></td>
</tr>
<tr>
<td>NDS</td>
<td>-.73**</td>
<td>.23</td>
<td></td>
<td>-.58**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIM</td>
<td>.01</td>
<td>.80**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDF</td>
<td></td>
<td></td>
<td>-.01</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**p < 0.01

In the current study, two scales corroborated the Bipolarity Hypothesis. Both PIM and DEF, measures of defensiveness, demonstrated very strong negative correlations with two scales containing rare symptoms (NIM and NDS). CDF behaved very differently from all other scales and demonstrated no significant correlations at all. It showed non-significant positive correlations with feigning indicators, but showed negligible correlations with other defensiveness indicators, PIM (.01) and DEF (-.01). Notably, CDF produced the smallest effect size (d = 0.94) of all Spanish PAI validity indicators when distinguishing between defensive and honest responders. CDF uses the scores of 6 different PAI scales to create a function score, so it is possible that it does not measure the same construct in the current sample than the English Version of the PAI. Besides the CDF, MAL did not support the bipolarity hypothesis because of its
strong negative correlation (−.56) with another feigning indicator (NDS) and non-significant correlations with PIM and DEF. Interestingly, MAL also demonstrated the lowest positive predictive power of all feigning indicators (see Table 9), indicating it was the least effective in correctly identifying malingers.

Effects of Clinical Symptoms on Validity Indicators

The supplementary question sought to investigate the relationship between patients’ primary diagnosis and their scores Spanish PAI validity scales. Separate analyses of variance (ANOVAs) were conducted for the general diagnostic groups of clinical disorders identified in patient charts (i.e., mood disorders and anxiety disorders), with the diagnostic category as the independent variable (IV) and Spanish PAI validity scale scores as the dependent variable (DV). Cohen's $d$ were computed to measure effect sizes.

Table 18

*Differences on the Spanish PAI Validity Indicators for Patients Diagnosed with Only Mood Disorders in the Honest Condition*

<table>
<thead>
<tr>
<th>PAI scales</th>
<th>Mood Disorder ($n = 19$)</th>
<th>Other Disorder ($n = 9$)</th>
<th>F</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>NIM</td>
<td>70.45</td>
<td>22.58</td>
<td>65.55</td>
<td>19.90</td>
</tr>
<tr>
<td>MAL</td>
<td>57.38</td>
<td>14.11</td>
<td>50.33</td>
<td>6.76</td>
</tr>
<tr>
<td>RDF</td>
<td>63.34</td>
<td>14.05</td>
<td>58.15</td>
<td>9.54</td>
</tr>
<tr>
<td>NDS</td>
<td>12.16</td>
<td>9.00</td>
<td>9.89</td>
<td>6.64</td>
</tr>
<tr>
<td>PIM</td>
<td>47.55</td>
<td>13.33</td>
<td>44.49</td>
<td>18.38</td>
</tr>
<tr>
<td>DEF</td>
<td>2.63</td>
<td>1.98</td>
<td>2.33</td>
<td>2.55</td>
</tr>
<tr>
<td>CDF</td>
<td>149.88</td>
<td>14.38</td>
<td>140.45</td>
<td>16.60</td>
</tr>
<tr>
<td>INF</td>
<td>60.13</td>
<td>11.66</td>
<td>57.81</td>
<td>12.15</td>
</tr>
<tr>
<td>ICN</td>
<td>67.68</td>
<td>12.41</td>
<td>59.66</td>
<td>8.28</td>
</tr>
</tbody>
</table>
These analyses were conducted to compare the scores of patients with a primary diagnosis of mood disorder to other patients in the honest condition. As seen in Table 18, there were no significant differences in mean scores between these two groups, largely due to the very small samples. The moderate to large effect sizes evidenced by CDF and ICN could indicate the need for additional research on the potential effects of depression. However, power in the current study is too low to draw conclusions regarding whether the presence of a mood disorder affects classification on Spanish PAI validity indicators.

Originally, it was also planned to investigate whether other clinical diagnoses (i.e., anxiety disorders) displayed a significant relationship to patients’ scores on validity indicators. However, due to limited sample size and the small number of participants with different diagnoses in the Honest condition, this analysis could not be conducted.
Psychologists and other mental health professionals are aware that most standardized assessment measures were developed for clients proficient in English and subsequently normed on samples comprised mainly of European American individuals. However, contemporary methods of psychological assessment in the United States are beginning to face unique challenges in a rapidly changing cultural landscape with increased diversity among the populations needing mental health interventions. Researchers have long emphasized that cut scores established for normative samples do not generalize to members of specific minority groups. They have called for different cut scores to use in the interpretation of diagnostic measures for psychopathology (Correa & Rogers, 2010).

The need for culturally appropriate cut scores is particularly pronounced for individuals whose primary language is Spanish because, when comparing the mean scores of Hispanic Americans and European Americans even on English versions of multiscale inventories, culturally specific response patterns emerge. Language plays an increasingly important role in test validity because there is a growing segment of the United States for whom traditional measures in the English language cannot be effectively used (Solano-Flores, Backhoff, & Contreras-Niño, 2009). To date, only a small number of Spanish-language measures are properly validated. These measures mainly include multiscale inventories whose English language versions are widely used in research and clinical practice. Particular examples include the Spanish Minnesota
Ethical guidelines from the American Psychological Association require that psychologists working with ethnically, linguistically, and culturally diverse populations should recognize these characteristics as important factors affecting a person’s experiences, attitudes, and psychological presentation (Bersoff, 2004; Weiss & Rosenfeld, 2012). Psychologists can easily conclude that culturally-related factors also have important effects on assessment results when evaluated by standardized testing measures. Specifically, interpretation of test results based solely on guidelines developed for mainstream American culture and cut scores contained in the test manuals can lead to biased results and incorrect classification of individuals from different cultural groups (Dana, 2005). For example, a consistent pattern emerges with African Americans averaging 2 to 3 T points higher than European Americans across PAI clinical scales, and with raw score differences of > 5 on SOM, ANX, PAR, and SCZ (Correa & Rogers, 2010). In the PAI manual, Morey (2007) provides separate T score conversions for African Americans so that cultural response style may be incorporated into test interpretation. On this point, researchers agree that assessment bias can be minimized when clinicians are well-informed about the populations they are testing, recognize limitations of their measures, and use culturally-specific measures to aid in their interpretation of assessment results (Dana, 2005). However, Morey (2007) continues to recommend the use of the standard norms to “maintain the test’s interpretive consistency across demographic groups” (p. 91).
This issue of diversity in assessment is especially important when considering an individual’s preferred language and using test translations, because a translated measure does not necessarily retain the psychometric properties of the original language version (APA, 1993). These psychometric properties of standardized assessment measures are likely to change when administered to individuals who are culturally different from the normative sample (Marin & Marin, 1991). Furthermore, individuals who are not tested in their preferred language can suffer a detachment effect (Bamford, 1991) and fail to adequately connect with the assessment questions or fully express their emotional and psychological issues. The detachment effect can result in poor communication about symptoms and less self-disclosure (Dana, 1995); however, it is often remedied when individuals are tested in their preferred language. For example, Guttfreund (1990) shows that bilingual Hispanic American patients who prefer to speak Spanish are more able to effectively express their emotions when tested in that preferred language rather than English.

Throughout recent years, different professional organizations have addressed issues of diversity and created guidelines and standards for addressing these issues within the realm of psychological testing. For example, the Standards for Educational and Psychological Testing from the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (AERA, APA, NCME, 1999) address language and diversity by specifying that any oral or written test also measures an examinee’s verbal skills. According to the Standards, the reliance on verbal abilities creates a particular concern for individuals whose primary language is not the original language of the test. These standards
conclude that “in such instances, test results may not reflect accurately the qualities and competencies intended to be measured” (AERA, APA, NCME, 1999, p. 91). On this point, translated tests can create test bias, the possibility for misdiagnosis, and the serious misinterpretation of test results (Dana, 1993).

Issues of test bias are magnified when translated versions of assessment measures are used in professional settings. The Test Translation and Adaptation Guidelines developed by the International Test Commission (ITC; Hambleton, 2001) specify that test developers must apply appropriate research methods and statistical techniques to establish the validity of each translated test for the new target population. Only tests that have been formally translated and subsequently validated as translated tests should be used in clinical practice (Hambleton, 2001). To date, the PAI has been translated and published in Spanish as well as English. For the Spanish PAI, clinicians must take into account a client’s language preference prior to beginning the assessment process. In cases where client is bilingual and expresses only a minor preference, practitioners might choose the English version due to its extensive validation. When a strong preference is expressed for Spanish, or English language abilities are limited, the Spanish PAI would be the most appropriate.

The paucity of well-researched Spanish language testing measures is clearly evident in many domains of psychological assessment which include, but are not limited to, response styles such as malingering and defensiveness. To date, there is only one study that investigates malingering and defensiveness on the Spanish PAI (Fernandez et al., 2008). Since Spanish PAI validity scales have not yet been investigated with Spanish-speaking clinical populations, the current study focuses on determining
reliability and validity. The current study also investigates acculturation and appropriate cut scores for the interpretation of the Spanish PAI when distinguishing malingering and defensiveness from honest responding.

The following section presents an overview regarding the current state of Spanish language assessment measures with an emphasis on their clinical utility with Hispanic Americans. Results specific to the Spanish PAI and the current study are also addressed.

**Culturally-Specific Response Patterns and Hispanic Americans**

The impact of culture on response style is evident even on English language versions of standardized assessment measures. For example, research on the MMPI-2 has consistently found significant “L” scale elevations among Hispanic Americans when compared to European Americans (Callahan, 1998; Campos, 1991). The L scale was developed to detect attempts by patients to present themselves in a favorable light (Hathaway & McKinley, 1989). Elevated patterns suggesting that Hispanic Americans distort their self-reports to appear less impaired are not confined to one measure. Studies looking at the PAI yield similar results. For example, Hopwood, Flato, Ambwani, Garland, and Morey (2009) found that Hispanic American participants scored higher than European Americans on all socially desirable response measures used in the study. On this same point, Romain (2000) found that more than 40% of the PAI protocols from Hispanic Americans were considered “invalid” based on the standard cut scores outlined in the PAI manual (Morey, 1991), as compared to 20% of the European
American profiles. As a contributing factor, Hispanic Americans had higher Positive Impression Management (PIM) scores when compared to European Americans.

Findings about impression management and socially desirable responding might lead practitioners to surmise that Hispanic Americans are largely reticent to disclose their psychological issues in the formal context of an evaluation and, perhaps, this is why no other diagnostic patterns are sometimes evident on the clinical scales of these particular assessment measures. Hesitation to disclose symptoms might reflect an issue in response style and interview behavior for this population rather than indicate an absence of symptoms (Correa & Rogers, 2010). However, other theories of Hispanic American response styles suggest a different explanation. For example, the phenomenon of *Extreme Response Style* suggests that individuals of certain cultures, particularly Hispanic and Mediterranean cultures, have a tendency to respond at *either* the extremely low or the extremely high end of the spectrum when given choices on Likert-type scales in the United States (Hui & Triandis, 1989). It is believed that these individuals consider extreme responses to be more sincere than a “conservative” response located in the middle of a Likert-type scale. The distinction is most evident for individuals within these two cultures in contrast to individuals of Asian cultures, who do tend to respond in the middle of the scale (Zax & Takahashi, 1967). Notably, the language of a test can magnify this cultural response style. In a study that administered the same items in two different languages to bilingual individuals, Gibbons, Zellner, and Rudek (1999) found that participants used more extreme ratings (both high and low) when responding in Spanish than in English. Contrary to research stating that Hispanic Americans tend to respond defensively to multiscale inventories, studies of Extreme
Response Styles suggest that extreme responding is possible in both directions (i.e., underreporting and overreporting).

Table 19 demonstrates the current sample’s distribution of endorsement across all items on the PAI’s 4-point Likert-type scale. The honest condition is of particular interest because, to an extent, extreme scores are to be expected in the experimental conditions.

Table 19

Percent of Endorsement for PAI Ratings across Experimental Conditions

<table>
<thead>
<tr>
<th>PAI Responses</th>
<th>Honest</th>
<th>Malingering</th>
<th>Defensive</th>
<th>Total Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>46.5%</td>
<td>27.3%</td>
<td>61.9%</td>
<td>45.4%</td>
</tr>
<tr>
<td>1</td>
<td>16.0%</td>
<td>18.9%</td>
<td>9.3%</td>
<td>14.8%</td>
</tr>
<tr>
<td>2</td>
<td>12.5%</td>
<td>17.4%</td>
<td>7.3%</td>
<td>12.5%</td>
</tr>
<tr>
<td>3</td>
<td>24.1%</td>
<td>34.9%</td>
<td>20.4%</td>
<td>26.7%</td>
</tr>
<tr>
<td>% of Extreme</td>
<td>70.6%</td>
<td>62.2%</td>
<td>82.3%</td>
<td>72.1%</td>
</tr>
</tbody>
</table>

Note. Extreme is the sum of “0” and “3” responses.

The honest group demonstrated a high percentage of symptom denial (46.5%), corroborating models of increased defensiveness among Hispanic American patients. Notably, however, complete endorsement of items accounted for nearly one quarter of PAI responses among honest participants (24.1%). Extreme responding became even more pronounced in the defensive condition (82.3% extreme responses). Theses finding indicate that, although symptom denial remains the most prevalent response, Extreme Response Style is still evident in the current sample, with responses in the middle of the Likert-type scale receiving relatively little endorsement.
The study by Romain (2000) also casts doubt on the assertion that defensiveness is the predominant response style for Hispanic Americans. Despite finding a higher PIM score for Hispanic Americans, Romain (2000) noted that both Hispanic and European Americans showed relatively little withholding or defensiveness as demonstrated by low mean PIM scores of 45.32 and 38.06 respectively. PAI research on cultural response styles is lacking, in general, and the normative samples included in the PAI manual create three major limitations in interpreting results for Hispanic American patients. First, ethnic differences for Hispanic Americans are explored in the test manual for the census-matched standardized sample but were not considered for the clinical sample. A second major limitation is the collapsing of all minority groups except African Americans into a single “other” group (Romain, 2000; Todd, 2004). The clinical standardization samples described in the more recent version of the PAI manual (Morey, 2007) are composed of 78.8% European Americans, 12.6% African Americans, and 8.6% “other” minority groups. Combining all minority groups into a single category does not allow for specific comparisons between groups and it implicitly makes the erroneous assumption that all minority groups are alike, except for African Americans. Thus, this grouping also creates a third major problem by masking minority differences. For instance, high scores for Hispanic Americans on a particular scale might be balanced by low scores from another culture (Correa & Rogers, 2010).

Published research conducted with clinical samples has not systematically attempted to identify differences in response patterns of ethnic minority populations. Greene (2000) points out that very little research has examined differences between Hispanic Americans and European Americans on both clinical and validity scales of the
MMPI-2. With most of the research having been conducted on undergraduate students with presumably low levels of psychopathology, Greene cautions against making general statements about the cultural response styles of Hispanic American patients on the MMPI-2, concluding that it is premature for this clinical population and that further research is necessary.

A recent study using the Spanish language PAI takes an important first step in evaluating malingering among Spanish-speaking populations. In a within-subjects design, Fernandez et al. (2008) used a non-clinical sample of bilingual individuals to assess the performance of PAI validity scales across both language versions. They found that the validity scales, generally, performed similarly in both language versions, with the NIM and PIM scales demonstrating the highest levels of equivalence. Results also indicated possible defensiveness within the sample, as individuals responding honestly exhibited a greater tendency to underreport symptoms on the Spanish version. However, these differences were small and only the difference between English and Spanish responses on the DEF index was statistically significant ($d = 0.38$). Still, the authors advise that their results should be interpreted with caution, as their sample of bilingual individuals is different than most samples of monolingual Spanish speakers in levels of acculturation and education.

Table 20 compares effect sizes for feigning between the current sample and Fernandez et al.’s (2008) sample of bilingual participants taking the Spanish PAI.
Table 20

A Comparison of Effect Sizes Between Honest and Feigning Conditions

<table>
<thead>
<tr>
<th>PAI feigning indicator</th>
<th>Hispanic American non-clinical sample&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Hispanic American clinical sample&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIM</td>
<td>4.17</td>
<td>1.19</td>
</tr>
<tr>
<td>MAL</td>
<td>2.05</td>
<td>0.90</td>
</tr>
<tr>
<td>RDF</td>
<td>1.60</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Notes. For feigning indicators, NIM = Negative Impression Scale; MAL = Malingering Index; RDF = Rogers Discriminant Function. <sup>a</sup>These values were obtained from Fernandez et al. (2008). <sup>b</sup>These values were obtained from the current sample.

Generally, effect sizes are much larger for feigning indicators in Fernandez et al.’s bilingual sample. NIM scores for the bilingual sample were particularly high for the feigning condition in the bilingual college sample ($M = 124.04; SD = 21.58$) compared to the monolingual clinical sample in the current study ($M = 97.44; SD = 26.10$). Lower endorsement of NIM items could be due to cultural and clinical differences between the samples. For example, Fernandez et al. (2008) had a sample of highly educated bilingual individuals, while participants in the current study averaged approximately 10 years of education, with 75% of individuals receiving no education in the United States. While Fernandez et al. (2008) did not measure level of acculturation; it is likely that their bilingual sample of university students also represents a higher level of acculturation than that of the current sample.

As a clinical sample, the current sample was likely more knowledgeable concerning genuine symptoms than college undergraduates. Methodological considerations, such as the selection of scenarios and instructions can impact results of feigning studies (Rogers, 2008). Specifically, Fernandez et al. (2008) instructed those
in their feigning condition to pretend they had recently been arrested for a crime. Participants were told to appear so mentally ill that they should not be held responsible for the crime and should, therefore, be found “Not Guilty By Reason of Insanity” at trial. In the current study, the experimental instructions about the scenario were designed to be more familiar and relatable to patients. The instructions asked participants to feign symptoms in order to gain entry into a highly desirable mental health treatment program. Additionally, the current study stressed that symptom presentation must be convincing and participants were encouraged to “fool the examiner” into believing their fabricated presentations. Instructions that stress the importance of convincing presentations are common in malingering research (Rogers, 2008). However, instructions with this caveat may have produced attenuated results when compared to a study that did not include this caution.

As noted (see Table 21), effect sizes for PIM and DEF in Fernandez et al. (2008) were more than double than in the current study. Particularly with NIM, the effect size ($d = 4.17$) is vastly higher than feigning research with clinical samples.

Comparisons between Fernandez et al. (2008) and the current study yielded much smaller effect sizes for defensiveness indicators. One possible interpretation is that defensiveness is a more consistent response style among Hispanic Americans, despite level of education and acculturation. Smaller differences in effect size could also be due to the nature of instructions for participants in the defensive conditions of both studies. Specifically, Fernandez et al. (2008) asked participants in their defensiveness condition to present themselves favorably in order to obtain a highly desirable job. In the current study, participants were asked to present themselves
favorably to obtain highly desirable treatment services. Both of these instructional sets are more easily followed than an insanity defense using a criminal scenario (i.e., Fernandez et al., 2008).

Table 21

<table>
<thead>
<tr>
<th>PAI Defensiveness Indicator</th>
<th>Fernandez et al.</th>
<th>Current Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$d$</td>
<td>$d$</td>
</tr>
<tr>
<td>PIM</td>
<td>1.93</td>
<td>1.47</td>
</tr>
<tr>
<td>DEF</td>
<td>1.74</td>
<td>1.68</td>
</tr>
<tr>
<td>CDF</td>
<td>0.24</td>
<td>0.94</td>
</tr>
</tbody>
</table>

*Note.* For defensiveness indicators, PIM = Positive Impression Management; DEF = Defensiveness Index; CDF = Cashel’s Discriminant Function.

It is unclear why CDF was the only defensiveness indicator to produce only a minimal effect size in the Fernandez et al. (2008) study. However, in the current study, CDF also produced the smallest effect size for of all Spanish PAI validity indicators with non-significant correlations with NIM and MAL. CDF uses the scores of 6 different PAI scales to create a discriminant function score; so, it is quite possible that this pattern of score varies by language and cultural diversity.

Given the lack of feigning research with Hispanic American populations, a primary goal of the current study was to provide comprehensive data on validity indicators of the Spanish PAI. The following section discusses utility of Spanish PAI validity indicators in distinguishing response styles, reliability of the Spanish PAI, and the effects of acculturation on response patterns for Hispanic Americans on the Spanish
Comparisons are also made between Hispanic American results in this study and the normative data for European Americans on the English language version of the PAI.

Classification Accuracy for the Spanish PAI Feigning Indicators

The PAI, like nearly all other self-report measures is vulnerable to dissimulation based on how the examinee responds to test items. This measure also focuses on two unlikely detection strategies for malingering: Rare Symptoms and Spurious Patterns (Rogers & Correa, 2008). For the detection of underreporting, the Spanish PAI indicators focus on measures of defensiveness and social desirability (Morey, 2007).

A brief review of PAI scoring interpretation is helpful before discussing classification accuracy of the Spanish PAI. The basic determination of feigning or defensiveness relies on calculating T scores and indexes to determine whether the scores exceed a determined cut score. When applied to the Spanish PAI, the overall classification rates were low for several cut scores suggested throughout the literature (Hawes & Boccaccini, 2009; Morey, 2007). Therefore, the current study focused on determining cut scores that minimized the number of false positives for a sample of primarily Spanish-speaking Hispanic Americans.

The effectiveness of cut scores suggested in English PAI studies, such as those included in the PAI manual (Morey, 2007), as well as those in a recent PAI meta-analysis by Hawes and Boccaccini (2009) were evaluated and adjusted to minimize false positives (e.g., NPP ≥ .95). As suggested by Rogers et al. (2012), cut scores were also utilized to rule-out feigning (i.e., likely genuine) and rule-in feigning (i.e., likely feigning). For feigning indicators, NIM ≥ 115T yielded a specificity and positive
predictive power of 1.0, which—by definition—is consistent across base rates. For the current study, the NIM scale, which employs a Rare Symptoms detection strategy, produced the most effective rule-in and rule-out criteria for scores \( \geq 115T \).

As Table 9 demonstrates, the optimal cut scores identified by Fernandez et al. (2008) did not generalize to the current research. Without a clinical sample, a much lower NIM (\( \geq 81T \)) was effective. However, when applied to outpatients, the sensitivity rate plummeted to a mere .64. Because Fernandez et al. (2008) had equally high sensitivity and specificity, their use of Overall Correct Classification was justified. In the current investigation, this focus led to too many false positives.

Despite the lower overall correct classification (OCC) rates, cut scores determined by Fernandez et al. (2008) were appropriate for determining “likely feigning” protocols for all feigning indicators tested, except RDF. RDF, which is a feigning indicator based on combinations of items from various scales, produced clear rule-in criteria for malingering at much higher scores than those suggested by other researchers (Fernandez et al., 2008; Hawes & Boccaccini, 2009). Scores for RDF in the current study only reliably revealed likely feigning protocols at scores greater than or equal to 90T.

Generally, rare symptoms detection strategies such as NIM and NDS, produced the highest overall classification rates for Hispanic American patients. However, classification accuracy improves dramatically when scores forming an indeterminate range around the suggested cut scores are removed. The changes that occur in the NIM scale when this group is removed are particularly salient. Specifically, well-defined NIM cut scores which exclude the “too-close-to-call” cases improved specificity to 1.00.
This was the most pronounced effect on optimal cut score upon removal of the indeterminate group. As Table 10 demonstrates NIM ≥ 115T is the best single-point indicator for individuals who are likely feigning (NPP = .89 and PPP = 1.0) at a base rate of 15%. With the indeterminate group removed (see Table 11), NIM ≥ 110T is equally effective as the single-point cut score of NIM ≥ 115T. These estimates of utility are lower than the values for Spanish SIRS-2 primary scales, where the overall classification rate was high at .88. For the Spanish SIRS-2, Sensitivity (.90) and specificity (.85) were well balanced (Correa & Rogers, 2010). Regarding the Spanish PAI in this study, however, Sensitivity was extremely low at NIM ≥ 115T (.29) and Specificity was high (1.00). While this indicates a low false-positive rate for the Spanish PAI, this is achieved at the expense of correctly identifying large portions of malingerers.

For honest responders, PIM ≥ 72T demonstrates a positive predictive power of 1.0, indicating that all outpatients classified as defensive were, in fact, instructed to alter their response style to artificially present themselves in the best possible light on the Spanish PAI. For clinicians seeking to accurately identify defensive participants, PIM appears to be the most reliable scale due to the generally poor performance of CDF and DEF. Specifically, CDF, which considers items from several different PAI scales produced no clear rule-in or rule-out cut scores for defensiveness. Moreover, the DEF cut scores were relatively ineffective at differentiating between likely genuine and likely defensive presentations. Again, exclusion of an indeterminate range enables the Spanish PAI to better identify individuals responding defensively. With this exclusion, Negative predictive power increased for PIM cut scores across all base rates. This
increase in NPP indicates an increase in the PAI’s accuracy when classifying honest responders.

Importantly, practitioners should note that cut scores, which identified “likely defensive” responders in this study, were much lower than scores identified by previous researchers (Fernandez et al., 2008; Hawes & Boccaccini, 2009). As Table 12 demonstrated, PIM scores ≥61T identify significant underreporting of symptoms. The prevalence of defensiveness among Hispanic American outpatients yields high scores on the PIM scale even for honest responders. Using the construct of defensiveness as it is typically defined in the normative sample, it follows that lower cut scores are necessary to identify Hispanic Americans who are not minimizing symptoms. However, this practice leads large numbers of PAI profiles to be classified as uninterpretable. For example, lower cut scores for defensiveness scores Hispanic American patients potentially illustrate why 40% of Romain’s (2000) sample was excluded from analysis for yielding “invalid” profiles due to PIM scores higher than the 70T suggested in the PAI manual. Clinicians must utilize discretion when determining profile validity of Hispanic American patients when they yield higher defensiveness scores than European American patients. Depending on the acculturation level of their patients, it may be more appropriate to adjust cut scores for these individuals when interpreting the Spanish PAI, and determine how defensiveness may be affecting the clinical presentation of each patient on an individual basis.

Bipolarity Hypothesis for Feigning and Defensiveness

Morey and Lanier (1998) provide corroboration for the bipolarity hypothesis in
their early PAI meta-analysis. They found that scores on the PAI defensiveness indicators PIM and DEF are positively correlated with each other and negatively correlated with the three PAI measures of feigning (i.e., NIM, MAL, and RDF). In support of the Bipolarity Hypothesis, other studies have also found that feigners exhibit lower scores on measures of defensiveness. For example, Graham, Watts, and Timbrook (1991) found suppressed scores on the MMPI-2’s K scale for both male ($M = 35.8T$) and female ($M = 32.7T$) feigners in a simulation design. In an MMPI-2 meta-analysis, Rogers, Sewell, Martin, and Vitacco (2003) also found that most feigners do not show elevations on K.

In the current study, only PIM and DEF clearly supported the Bipolarity Hypothesis, demonstrating strong negative correlations with NDS and NIM. These two indicators also demonstrated relationships in the Morey and Lanier (1998) meta-analysis. Such findings support the Bipolarity Hypothesis, in part, indicating individuals who score high in defensiveness on some scales do tend to achieve low scores on scales containing rare symptoms.

Conversely, MAL only partially supported the bipolarity hypothesis in the current study. The MAL index showed a strong positive correlation with one feigning indicator (NIM) and a strong negative correlation with another feigning indicator (NDS). Interestingly, MAL also demonstrated the lowest positive predictive power of all feigning indicators, signifying it was the least effective in correctly identifying malingers.

Of the validity indicators, CDF behaved very differently from all other validity scales and indicators; it demonstrated no significant correlations at all. Unexpectedly, it showed non-significant positive correlations with feigning indicators, but negligible
correlations with other defensiveness indicators, PIM (.01) and DEF (-.01). Notably, CDF produced the smallest effect size \((d = 0.94)\) of all Spanish PAI validity indicators when distinguishing between defensive and honest responders in the current study. Because CDF uses the scores of 6 different PAI scales to create a function score, it is possible that it does not measure the same construct in the current sample than the English version of the PAI.

Reliability of the Spanish PAI

For measures of malingering, the English language version of the SIRS is considered the gold standard because of its exceptional reliability, validity, and classification accuracy (Blau, 1998; Lally, 2003). A study on the Spanish SIRS-2 found high reliability, validity, and classification accuracy for the adapted measure (Correa & Rogers, 2010). Comparable to the English version, whose primary scales exhibited high alpha coefficients \((M = .86; \text{range from } .77 \text{ to } .92)\) the alpha coefficients for the Spanish SIRS-2 were also generally high \((M = .89; \text{range from } .76 \text{ to } .96)\). The strongest alpha coefficients were found in scales that utilize amplified detection strategies: BL \((\alpha = .96)\) and SU \((\alpha = .95; \text{Correa, 2010})\). According to Rogers et al. (1992), these two primary scales also exhibited the highest alphas in the original English validation sample \((\text{BL } \alpha = .92; \text{ SU } \alpha = .92)\).

For the Spanish PAI, the internal consistency of each validity scale was moderate \((\alpha = .76 \text{ to } .78)\). With inter-item correlations in the acceptable range, these alphas indicate scale homogeneity.
Table 22

*A Comparison of Internal Consistency Determined by Alpha Coefficients Across English and Spanish PAI Studies*

<table>
<thead>
<tr>
<th>PAI Scale</th>
<th>English PAI</th>
<th>Spanish PAI</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIM</td>
<td>-</td>
<td>.72</td>
</tr>
<tr>
<td>NIM</td>
<td>.76</td>
<td>.71</td>
</tr>
<tr>
<td>NDS</td>
<td>.74</td>
<td>-</td>
</tr>
</tbody>
</table>

*Notes.* For validity scales, PIM = Positive Impression Management; NIM = Negative Impression Management; NDS = Negative Distortion Scale. Only alpha values that were published in each study are included in this table.

The current alpha levels are close to those found in existing Spanish and English PAI literature, even when comparing Hispanic American and European American samples (Mogge et al., 2010; Morey, 2007). However, NIM’s internal consistency was much lower in an earlier study of bilingual Hispanic American outpatients being administered the Spanish PAI (Rogers & Flores, 1995). Notably, Rogers and Flores administered Spanish language versions of the PAI to both monolingual and bilingual participants. Commonsensically, bilingual participants likely have higher levels of acculturation than monolingual Spanish-speakers and the participants in the current study. Rogers and Flores (1995) did not test for acculturation within their sample, but differences in cultural response patterns attributable to acculturation could have lowered internal consistency in their PAI scales.
Validity of the Spanish PAI for Feigning Indicators

Large effect sizes are crucial for establishing the discriminant validity of the Spanish PAI between feigning and genuine groups. Results from this simulation design indicate that the Spanish PAI produced moderate to very large effect sizes across all feigning indicators ($M_d = 1.04$; range from 0.90 to 1.35). Notably, effect sizes for validity indicators of the Spanish PAI are comparable to effect sizes noted for English language measures with detection strategies for the assessment of feigning: the MMPI-2 ($M_d = 1.31$), and the original PAI ($M_d = 1.45$; Jackson et al., 2005; Rogers, 2008; Rogers et al., 2003).

To date, the only Spanish language measure of feigning is the Spanish SIRS-2. Direct comparisons can be made between effect sizes from the Spanish PAI and the Spanish SIRS-2. The Spanish SIRS-2 produced very large overall effect sizes when distinguishing feigners from honest responders ($M_d = 2.00$; Correa & Rogers, 2010). Overall, Spanish SIRS-2 scales using amplified detection strategies (i.e., BL, SU, SEL, and SEV) produced somewhat higher effect sizes ($M_d = 2.19$ versus $M_d = 1.80$) than those utilizing unlikely detection strategies (RS, SC, IA, and RO) for Spanish-speaking Hispanic Americans. Amplified detection strategies also showed relatively higher effect sizes ($M_d = 1.90$) in the original validation sample than unlikely detection strategies ($M_d = 1.57$). This finding is of particular importance regarding the Spanish PAI because the PAI primarily uses the rare symptoms strategy (an unlikely detection strategy) to detect feigning (Morey, 2007).

The Spanish PAI can also be compared to the MMPI-2, which also has validity scales. In a mixed sample of clinical and non-clinical Spanish-speaking adolescents in
Mexico, Lucio, Duran, Graham, and Ben-Porath (2002) found that four scales (F, F1, and F2 scales, and F-K index) on the Mexican version of the *The Minnesota Multiphasic Personality Inventory-Adolescent* (MMPI-A; Lucio, 1998) adequately discriminated between feigners and honest responders. However, the authors generally found that higher cut scores were necessary in their sample of adolescents in Mexico. Thus, the authors caution against applying the findings from their study to Hispanic adolescents from the United States, highlighting that cultural differences between adolescents in Mexico and Hispanic American in the United States require different cut scores. Specifically, Lucio, et al. (2002) state that different cut scores might be because they have noted that Hispanic American adolescents in the United States tend to be less forthcoming when reporting symptoms than adolescents in Mexico.

The current investigation included comparisons with previous research results using the Spanish PAI, both within and between cultures. For the former, cultural differences were explored by considering participants on the basis of their ARSMA-II level of cultural identification. Efforts to assess cultural differences were only partially successful because most of the current sample had a Traditional orientation according to the ARSMA-II, indicating little cultural heterogeneity among participants. High levels of cultural homogeneity are expected in a sample of primarily Spanish-speaking participants. For the latter, the Hispanic American sample in this study was also contrasted with the original normative sample for the English language version of the PAI.

Of the three PAI validity scales, Negative Impression Management (NIM) is most often used to assess malingering. A meta-analysis by Hawes and Boccaccini (2009)
found the NIM scale for the English version of the PAI consistently produced the largest effect sizes when compared to MAL and RDF for detecting malingerers across studies. In the current study, the largest effect size was produced by NDS ($d = 1.35$), which was recently found to demonstrate a much higher effect size than other feigning indicators for the English version of the PAI (Rogers et al., 2012).

Differences in the average effect size across measures of amplified detection strategies between primarily Spanish-speaking Hispanic Americans and English language validation samples could be partly due to cultural factors. Findings indicate that Hispanic American individuals may have more difficulty identifying symptoms that European American individuals consider to be uncommon or unlikely, making them less prone to endorse these items when attempting to malinger (Correa & Rogers, 2010). Alternatively, smaller effect sizes for unlikely detection strategies, particularly on the Spanish SIRS-2, could reflect defensiveness—even in the feigning condition. It could also reflect a reticence to endorse symptoms of extreme pathology, even when attempting to feign complete impairment. In either case, amplified detection strategies are more effective for this population.

An unexpected finding relating to response style, the INF scale produced a larger effect size ($d = 1.23$) than NIM in the current study. INF was designed to detect inconsistent responding by individuals who do not yield valid PAI protocols for reasons such as carelessness, confusion, or reading difficulties (Morey, 2007). Traditionally, INF scores are not used to detect potential malingering. However, the significant differences between honest and feigning conditions in the current sample indicates the possibility of idiosyncratic interpretations of its item content for the Spanish PAI. Of
particular note, INF Item 40 (“My favorite poet is Raymond Kertezc.”) shows a notable discrepancy between the honest and defensive conditions in the current study. Interestingly, no honest responders endorsed the item. The mean scores for the malingering and defensive conditions were $M = .72$ and $M = 1.04$, respectively. A much higher discrepancy was noted on Item 320 (“In my free time I might read, watch TV, or just relax”). Item 320 demonstrated a notably higher average score among participants in the malingering condition ($M = 2.12$) than for participants in both the honest ($M = .35$) and defensive ($M = 1.04$) conditions. This discrepancy indicates the possibility of cultural bias regarding the perception of persons who engage in these behaviors.

Effects of Acculturation on the Spanish PAI

In psychological assessment, issues of acculturation must be considered for individuals whose primary identification is toward a different culture (i.e., the traditional orientation, as classified by the ARSMA-II). Researchers and practitioners both recognize that standardized assessment measures administered to individuals who are culturally different from the normative sample can have quite different psychometric characteristics and lead to biased results as well as incorrect classification of individuals from different cultural groups (Marin & Marin, 1991; Dana, 2005). In order to avoid inappropriately making generalizations about different cultural identifications among participants in the current sample, this study evaluated possible effects of acculturation on the Spanish PAI. This practice is advisable because English language measures adapted for Spanish speakers frequently fail to evaluate level of acculturation (Echemendia & Harris, 2004; Salazar, Perez-Garcia, & Puente, 2007; Renteria et al,
By comparing their utility estimates and optimal cut scores to adolescent samples from the United States, Lucio et al. (2002) point out the detrimental effects of failing to acknowledge cultural differences in their study of the MMPI-A and call for different cut scores when the same measure is used for adolescents in Mexico and American adolescents of Hispanic descent.

The current study attempted to analyze correlations between level of acculturation and performance on Spanish PAI validity indicators to determine if a relationship existed between scale scores and levels of acculturation. The only significant relationship between validity indicators and ARSMA-II Acculturation Score proved to be a small positive correlation with MAL. The absence of significant relationships between acculturation and validity indicator scores could denote that acculturation is not a valid predictor of response style on the Spanish PAI. However, it should be noted that the absence of a significant relationship is likely due to the cultural homogeneity of the present sample. Since the majority of the current sample was classified as having a “traditional” orientation, study results do not generalize to Hispanic Americans, who are classified as bicultural or assimilated according to the ARSMA-II. The only published Spanish PAI feigning study was conducted with bilingual individuals, who likely have a vastly different level of acculturation from participants in the current study (Fernandez et al., 2008). However, cultural heterogeneity of samples from previous research studies cannot be inferred because all existing research has neglected to study level of acculturation.
Effects of Psychopathology on Spanish PAI Classification

The current study examined whether validity indicators are affected by Axis I diagnoses. The rationale behind investigating these diagnostic differences is that patients with genuine disorders (e.g., schizophrenia and PTSD) sometimes have elevated scores on the MMPI-2 (Rogers et al., 2003).

To date, the effects of Hispanic culture on the clinical scales of multi-scale inventories such as the MMPI-II and PAI has not been researched (Correa & Rogers, 2010). The lack of research in this area is likely because high scores on defensiveness indicators among Hispanic Americans render clinical protocols uninterpretable due to underreporting of symptoms (Correa & Rogers, 2010; Romain, 2000). Distinct patterns of Axis I symptomatology emerge for other cultural groups. For example, African-Americans tend to endorse more symptoms of paranoia, without necessarily suffering from clinically significant psychopathology (Correa & Rogers, 2010; Todd, 2005). However, no such patterns have been discovered for Hispanic Americans. Lower rates of general symptom endorsement among Hispanic Americans has likely precluded researchers from discovering culturally-influenced response patterns on PAI clinical scales.

A patient’s diagnosis can often affect elevations on validity indicators. In a meta-analysis of the MMPI-2 and malingering, Rogers, Sewell, Martin and Vitacco (2003) reviewed detection strategies. One main focus of the MMPI-2 is “quasi-rare” strategies such as those found on the F and Fb scales. The term “quasi-rare” signifies that the items are uncommon within normative samples, but not necessarily among genuine clinical patients. Rogers and Bender (2003) cautioned against relying exclusively on F-
scale elevations because true patients with severe psychotic disorders might be misclassified. Specifically, a high score on the F-scale is not necessarily indicative of malingering; instead, it can mean that the person is responding honestly and exhibits genuine, albeit uncommon, symptoms such as those found in schizophrenia.

The PAI NIM scale employs a rare symptoms detection strategy, so it can be inferred that the scale is also susceptible to elevation from genuine patients reporting symptoms. In the current sample, there was not a sufficient number of patients with psychotic disorders (potentially rare symptoms) for analysis within the honest condition. However, patients with a primary diagnosis of mood disorders were studied to determine whether their scores on validity indicators were different from the rest of the sample. There were no significant differences in mean scores when patients with a primary diagnosis of depression or bipolar disorder were compared other patients in the honest condition. However, power in this study is far too low to draw conclusions regarding whether the presence of a mood disorder affects classification on Spanish PAI validity indicators. Therefore, this study cannot determine whether the presence of a mood disorder affects classification on Spanish PAI validity indicators.

A second proposed analysis could not be conducted because there were no honest participants that were only diagnosed with anxiety disorders. An attempt was made to modify this analysis and compare group differences among (1) patients who were diagnosed with both an anxiety and mood disorder and (2) other honest participants. Again, there were no significant differences in feigning indicators. However, power was much lower than in the comparison described above, as only seven participants in the honest condition were diagnosed with an anxiety disorder.
Implications for Professional Practice Using the Spanish PAI

In line with the ITC test guidelines, test translations should not be used for clinical evaluation until validated for their intended purpose and target population (Hambleton, 2001). The Spanish PAI was created using a back-translation procedure recommended by most researchers (Matias-Carrelo et al., 2003; Marin & Marin, 1991). The current study sought to examine its accuracy in distinguishing between honest, defensive, and feigning response styles in the assessment of a Spanish-speaking Hispanic American clinical population.

Throughout different domains of psychological assessment, few Spanish language measures have been adequately researched and validated for use with Spanish-speaking Hispanic American populations. Studies of Spanish-language multi-scale inventories with embedded validity scales (i.e., MMPI-2 and PAI) have, thus far, neglected to include analyses of these validity scales and associated response styles such as malingering and defensiveness in adult clinical populations (Correa & Rogers, 2010; Fernandez et al., 2008; Lucio et al., 2002; Romain, 2000). Because the classification of malingering and defensiveness often has important implications for how clinical patients are treated (Rogers & Schuman, 2005), the current study sought to provide data on the utility of the PAI validity indicators for Spanish-speaking populations.

Results from the current study and past research using the Spanish PAI (Fernandez et al., 2008) indicate the Spanish PAI can be a useful and valid measure for the classification of malingering and defensiveness, when using different cut scores than those traditionally used by clinicians based on European-American normative
samples (Morey, 2007). However, clinicians should exercise great care in choosing appropriate cut scores for their patients, as studies have identified different optimal cut scores based on acculturation, education level, and other demographic variables inherent in their samples.

Psychologists conducting assessments with the Spanish PAI should weigh several recommendations highlighted in multicultural assessment literature. Assessment bias is minimized when clinicians are well-informed about the populations they are testing, recognize limitations of their measures, and use culturally-specific measures to aid in their interpretation of assessment results (Dana, 2005). Therefore, depending on the level of acculturation of any particular client, clinicians may wish to consider using the different cut scores suggested by this study or the pre-existing literature (Fernandez et al., 2008). Conversely, clinicians may choose to follow Morey’s (2007) recommendation to use the standard norms and “maintain the test’s interpretive consistency across demographic groups” (p. 91). To reconcile these two disparate practice recommendations, clinicians may, instead, wish to include cautionary statements for all PAI interpretations involving clients with low levels of acculturation (Correa & Rogers, 2010). Utilizing this approach, a clinician can explain the implications of using different cut scores for the Spanish PAI and clarify the reasons for doing so, based on data from other tests used in the evaluation.

Practitioners should maintain awareness that elevations on validity scales for Hispanic American patients may reflect a specific response style (e.g., malingering or defensiveness), or it may reflect ethnically sensitive content. To properly convey these
alternatives in the results of an evaluation, clinicians should explicitly address both possibilities in a clinical report (Correa & Rogers, 2010).

Limitations of the Current Study

The current study contributes significantly to the literature on the Spanish PAI with its particular focus on the use of response styles, notably feigned mental disorders. Because of its intentional focus on primarily Spanish-speaking outpatients, it is not surprising that there was very little variability in level of acculturation among participants. This homogeneity limited the extent to which the relationship between acculturation and response style could be studied. Future research with a more culturally diverse sample of Hispanic Americans can shed light on this area (Salazar et al., 2007).

A second limitation observed in the current study was the lack of variability in diagnoses as well as overlap in diagnoses. The vast majority of the sample warranted diagnoses of both mood and anxiety disorders. Diagnostic comorbidity is common in clinical populations, especially between depression and anxiety (Almeida, Draper, Pirkis, Snowdon, Lautenschlager, Byrne, & Pfaff, 2012). Consistent with past research (Correa, 2010), psychotic disorders were under-represented in the current investigation.

Another factor limiting the present ability to assess whether genuine clinical symptoms affect validity scores, was the small size of the sample assigned to the honest condition. While the number of participants in the honest condition allowed sufficient statistical power for the primary analyses, important aspects of the supplementary question could not be addressed. Specifically, there were no
participants with anxiety disorders that did not also have diagnosed mood disorders. Additionally, there were only three individuals with psychotic symptoms in the honest condition and this small number did not allow for analysis of whether genuine endorsement of these symptoms would have affected feigning indicators which utilize rare symptoms strategies. Since this analysis could not be conducted, this study is unable to address how Spanish PAI feigning scales might be affected compared to Spanish MMPI-2 feigning scales. A study with a much larger sample size would allow for thorough investigation of diagnostic categories and their impact on validity indicators for honest responders. Ideally, the impact of anxiety symptoms would be investigated for individuals who do not have comorbid depression. Also, the impact of psychotic symptoms on feigning indicators would be explored.

A final important limitation was the use of only one measure to evaluate response styles. Multiple measures (e.g., MMPI-2 and PAI) would have allowed more systematic analyses of clinical symptoms and response styles. Use of the PAI and a structured interview such as the Spanish SIRS-2 would have allowed for a multi-method approach of studying response styles among Spanish-speaking patients.

Future Directions

Language equivalence could not be tested in the current study, because the sample was largely monolingual. Thus, no direct comparisons can be made about the Spanish and English language versions of the PAI. To date, the only published literature on the Spanish PAI validity scales has found very similar scores between both versions for bilingual participants (Fernandez et al., 2008). However, the Fernandez et
al. (2008) study was conducted with a non-clinical population whose level of education was notably higher than that of the current sample. ITC guidelines recommend language equivalence testing as part of the test adaptation process. However, this research has been focused on (a) non-clinical populations (Fernandez et al., 2008), (b) clinical scales, to the exclusion of validity scales (Fantoni-Salvador & Rogers, 1997), and (c) has not yet addressed the effects of acculturation differences on language equivalence (Fantoni-Salvador & Rogers, 1997; Fernandez et al., 2008; Rogers & Flores, 1995).

Culturally-specific response patterns for Hispanic Americans on multi-scale inventories have been vastly under-researched, to date. It is hypothesized that heightened levels of defensiveness tend to attenuate any notable patterns on clinical scales, due to general under-reporting of symptoms. It is important that future studies attempt to study potential patterns to aid with test interpretation for Hispanic American clients. Given the large number of protocols that have been deemed “invalid” and “uninterpretable” due to high PIM scores, it is advisable for researchers to refine the scales to minimize cultural effects, rather than excluding high PIM cases from analysis (Correa & Rogers, 2010; Romain, 2000).

While it is important to study the effects of acculturation, this construct can potentially be confounded by other demographic variables. For example, future studies should be conducted with clinical populations of Hispanic Americans with higher levels of education and these studies must encompass a broader range of education, as well. Future studies with more varied levels of acculturation, as well as higher variance in levels of education could help determine whether the current findings are primarily due
to cultural variables or whether they are more closely related to socioeconomic factors. These studies could also help identify factors that moderate and mediate the relationship between culture and Spanish PAI score.

Further research should also be conducted on the cultural content of the INF scale. The current study revealed a possible cultural bias in the endorsement of at least two items (item 40 and item 320). Without having conducted an investigation of participant’s opinions regarding the content of these items, however, it is impossible to determine the source of the potential test bias or how it may be remedied. INF scores and ICN scores were notably high in this sample. They have not been previously studied in Spanish PAI research, so future studies should assess the applicability of these scales to Hispanic Americans with Traditional cultural orientations.

Lastly, the only published research that currently exists on Spanish PAI validity indicators involves simulation studies. While simulation studies are informative, studies with known-groups comparisons of suspected malingerers in clinical settings are crucial for the validation of response style measures (Rogers, 2008). A next reasonable step in studying malingering would be to study these constructs in known malingerers, as well as other clinical and forensic populations.
Identification number: ____________

Age: __________

Gender: __________

Your Country of Origin: ________________________________

Father’s Country of Origin: ________________________________

Mother’s Country of Origin: ________________________________

Age you moved to the US: ________________________________

SES:  Lower  Middle  Upper

Education Completed:  Elementary ____  Middle ____  High School ____

                      GED  Vo. Tech  Some College ____

                      Bachelors  Masters  Doctorate

Years of Education Completed in the United States: ________ Grades: ____ to ____

Years of Education Completed in __________ (country): ________ Grades: ____ to ____

Employment History:  Unemployed  Part-time  Full-time

(Last Year)

                       Disabled

First Language Spoken: ________________________________

Other Languages Spoken: ________________________________
APPENDIX B

ROLE-PLAYING INSTRUCTIONS A: GETTING THE BEST TREATMENT FOR YOU AND YOUR FAMILY
Your Goal:

For this study, please role-play a person who is pretending to be completely disabled by their mental illness. Assume there is a new program at the treatment center. Patients who are a part of this program will receive:

- Free psychological services
- Free medical services
- Free services for family members
- First choice of appointment times
- Free transportation to and from the treatment center
- Priority over other patients in the waiting area and will be attended first when arriving for appointments
- Priority in signing up for workshops and groups at the center

Admission to this new program is only for the patients who are identified as the most severely impaired on this test. Once you are admitted to the program, you will continue to receive these benefits for the next two years, no matter how your symptoms may change later. You have to convince the person administering the test that you have a severe mental disorder that qualifies you as one of the sickest patients at this center.

Although this is only for a research experiment, please try to be as convincing as possible. It may sound easy, but the hard part will be really convincing the interviewer that you are so incapacitated by the symptoms that you are pretending to have.
Your Reward:

Can you fool the examiner? These tests are made to catch people who are trying to fake a mental disorder. Are you clever and convincing enough to avoid getting caught? You will receive $10.00 for being successful.

Before beginning the study, please take a moment to think about how you will answer the questions to appear so sick. You will be asked about this later.
Role-Playing Instructions B: Getting the Best Treatment for You and Your Family

Your Goal:

For this study, please role-play a person who is pretending to portray themselves as well adjusted. Assume there is a new program at the treatment center. Patients who are a part of this program will receive:

- Free psychological services
- Free medical services
- Free services for family members
- First choice of appointment times
- Free transportation to and from the treatment center
- Priority over other patients in the waiting area and will be attended first when arriving for appointments
- Priority in signing up for workshops and groups at the center

Admission to this new program is only for the patients who are identified by this test as not having very many problems. Once you are admitted to the program, you will continue to receive these benefits for the next two years, no matter how your symptoms may change later. You have to convince the person administering the test that you do not have very many symptoms or problems and you qualify you as one of the least sick patients at this center.
Although this is only for a research experiment, please try to be as convincing as possible. It may sound easy, but the hard part will be really convincing the interviewer that you are relatively unaffected by any symptoms or problems.

**Your Reward:**

Can you fool the examiner? These tests are made to catch people who are trying to minimize their symptoms on purpose. Are you clever and convincing enough to avoid getting caught? You will receive $10.00 for being successful.

Before beginning the study, please take a moment to think about how you will answer the questions to appear able to handle life’s difficulties. You will be asked about this later.
**Accurate Presentation of Symptoms:**

**Your goal:**

Please be open and honest in describing your symptoms and circumstances. Your job is to provide an accurate presentation of your current symptoms and psychological problems. You will receive $10.00 for your honest participation.

**Importance:**

Please take this study seriously. There are not many psychological tests available for people who speak Spanish. Your participation will help us make sure this Spanish language test is useful and accurate when it is used.
APPENDIX C

MANIPULATION CHECK AND DEBRIEFING
1. The study you just participated in asked you to follow the instructions you were given. Please briefly describe what your instructions asked you to do. [record verbatim] ___ correct, ___ incorrect

2. What situation were you asked to pretend you were in?

3. Did you follow the instructions?
   Yes   No

4. How hard did you try to follow the instructions?
   Didn’t try hard, it’s just a study ______
   Tried a little bit ______
   Gave a medium effort ______
   A good effort, I tried hard ______
   Excellent effort, I really tried to do my best ______

5. Were you comfortable participating in this activity?
6. Were you aware that there were questions designed to see if you were faking?

7. How do you think these questions were supposed to work? [record verbatim]

8. [Malingering and defensive conditions only] Do you think you were successful at deceiving the tests?
   Yes  No

9. [Malingering condition only] When faking, did you have a particular disorder in mind?
   Yes  No
   If yes, what was it?
REFERENCES


