Internal Consistency and Convergent Validity of the Personality Assessment Inventory English and European-Spanish Version with English/Spanish Bilinguals

Lauren E. Kois[1], Crystalann Rodriguez[2], Preeti Chauhan[3], Jessica Pearson[4] and Virginia Barber Rioja[5]

[1] The University of Alabama (USA)
[2] New York University (USA)
[3] The City University of New York (USA)
[4] Elmhurst Hospital Center (USA)
[5] Correctional Health Services NYC Health + Hospitals (USA)

Abstract. Practitioners have few personality inventory options when assessing Spanish-speakers, despite professional guidelines that encourage them to administer measures that are validated with their populations of interest. To build on research in this area, we examined the internal consistency and convergent validity of the Personality Assessment Inventory (PAI) and the Personality Assessment Inventory European-Spanish (PAIE-S) version among bilingual Latin American Spanish-speakers (final *n* = 53). For the PAI, 72.72% of scales and 35.48% of subscales had alphas above .70. For the PAIE-S, 50.00% of scales and 25.81% of scales met this alpha. Participants tended to score lowest on the PAI Alcohol Problems scale (*T* = 47.19) and the PAIE-S Warmth scale (*T* = 45.49). On average, participants scored highest on the PAI's Paranoia-Hypervigilance scale (*T* = 61.15) and the PAIE-S's Paranoia scale (*T* = 57.64). We identified 10 scales and subscales on which participants were significantly more likely (*p* < .00094) to score higher on one measure than the other. Participants more often scored higher on the PAI than the PAIE-S. All parallel scales and subscales converged at *p* < .00094 with the exception of the Antisocial Features-Egocentricity scale. Taken together, findings suggest taking caution when administering these measures to Latin American bilingual Spanish-speakers.

Estimates indicate that approximately 527,976,150 individuals speak Spanish worldwide (Simons & Fenning, 2017). Increasing numbers of Latin Americans are immigrating to the United States and 37 million individuals residing in the United States speak Spanish at home (Flores, 2017). These numbers are increasingly relevant for mental health practitioners whose psychological evaluations may provide diagnostic clarification and guide treatment planning for Spanish-speaking clients. At present, the psychometric properties of only a few personality measures have been explored with English/Spanish bilinguals.

Personality Assessment of Spanish-speakers

American Psychological Association (APA) guidelines indicate that practitioners are expected to individualize test selection and administration according to clients' unique characteristics (American Psychological Association, 2017). For example, individuals who primarily speak or prefer Spanish should be administered assessments written in and validated with Spanish-speakers. Fernandez, Boccaccini, and Noland (2007) recommend that practitioners should consider which translated tests are available, identify relevant research regarding the tests' administration and interpretation, evaluate the research relevant to the clients' background, and assess their own level of confidence that

the tests are appropriate for use with specific clients. In the United States, these guidelines are difficult to follow in practice because practitioners have limited options regarding multiscale inventories that are translated and validated with Spanish-speakers. Further, there is only a small body of research to reference on this topic (Fernandez et al., 2007; Weiss & Rosenfeld, 2012).

Personality and clinical assessment research show some differences in score profiles between non-Hispanic and Hispanic populations (Estrada & Smith, 2017; Gurven, von Rueden, Massenkoff, Kaplan, & Lero Vie, 2013). Furthermore, researchers have found significant differences in personality assessment protocols across Latin Americans by country of origin. For instance, Fantoni-Salvador and Rogers (1997) identified significant differences in clinical scales using the Personality Assessment Inventory Spanish Edition (PAISE; Morey, 1991–2007) among Latin American, Mexican American, and Puerto Rican Spanish-speakers. This suggests that differences in country of origin could play a role in personality profiles as measured by standardized assessments. Others (Boscán et al., 2000) found significant differences in Minnesota Multiphasic Personality Inventory–2 (MMPI–2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) Spanish translation protocols according to country of origin in their sample of Mexican and Venezuelan college students and Colombian college students and community members. Thus, research indicates that country of origin in the context of personality assessment should be further explored.

The Personality Assessment Inventory (PAI; Morey, 1991–2007) is a popular personality assessment among U.S. practitioners (Wright et al., 2017). Collectively, its 22 scales are meant to assess a broad range of characteristics. It contains four validity scales (Inconsistency, Infrequency, Negative Impression, Positive Impression), which measure response style. The 11 clinical scales are intended to assess symptoms of psychopathology: Somatic Complaints, Anxiety, Anxiety-Related Disorders, Depression, Mania, Paranoia, Schizophrenia, Borderline Features, Antisocial Features, Alcohol Problems, and Drug Problems. Five treatment consideration scales focus on Aggression, Suicidal Ideation, Stress, Nonsupport, and Treatment Rejection. Two interpersonal functioning scales aim to assess Dominance and Warmth. Furthermore, there are 28 subscales within the clinical scales and three subscales within the treatment consideration Aggression scale. The interested reader may reference Morey's PAI manual or Morey (2003) for detailed scale and subscale descriptions.

The PAI has a number of strengths: a relatively low (fourth grade) reading level requirement, brief administration time, and low long-term administration costs.

It has demonstrated utility in a range of clinical contexts, such as providing diagnostic clarification (Edens & Ruiz, 2008); differentiating levels of care (Sinclair et al., 2015); and forensic referrals, such as assessing risk for violence (Gardner, Boccaccini, Bitting, & Edens, 2015). However, practitioners should not assume that the mounting research on the PAI with predominantly U.S. born, English-speaking samples generalizes to Spanish-speakers. Below we summarize research on the two available Spanish versions of the PAI–the Personality Assessment Inventory-Spanish Edition (PAISE; Morey, 1992) and the Personality Assessment Inventory European-Spanish version (PAIE-S; Ortiz-Tallo, Santamaría, Cardenal, & Sánchez, 2011).

Personality Assessment Inventory Spanish Edition (PAISE) Research

The PAISE is a direct translation of the PAI from English to Spanish and it has yet to be validated with a Spanish-speaking population, an important step per International Test Commission guidelines (American Education Research Association, American Psychological Association, & National Counsel on Measurement in Education, 2014). There are some complications with this translation. First, direct English-to-Spanish test translation may not be a valid method for facilitating the psychodiagnostic assessment of English/Spanish bilinguals. English assessments that are validated with U.S. English-speakers may successfully measure constructs among this population. However, when directly translated and applied in other languages, test items may take on different meaning and tone, thereby creating confounds that may interfere with test reliability, validity, and its ultimate utility. Second, language, culture, and ethnicity may also influence psychometric properties. Individuals who share the same language do not necessarily share the same culture (Sue & Sue, 2012). Third, differences across Hispanic subcultures may impact PAI and PAISE measurement outcomes across these groups. Therefore, it is critical that researchers explore the PAI and its iterations among Hispanics and English/Spanish bilinguals in order to understand their generalizability across subcultures.

Three studies have explored aspects of the PAISE's psychometric properties in the United States. Rogers, Flores, Ustad, and Sewell (1995) administered the PAISE to a sample of Mexican American Spanish-speakers (three quarter monolingual) engaged in substance abuse treatment. They found that bilinguals' PAISE and PAI protocols had adequate convergence. The PAISE had satisfactory test-retest reliability but low average scale internal consistency ($\alpha = .63$). In addition, three of four PAISE validity scales had

poor internal consistency and protocols were invalid twice as often as normative PAI protocols. Conversely, Fernandez, Boccaccini, and Noland (2008) administered the PAI and the PAISE to English/Spanish bilinguals and found that Negative Impression Management and Positive Impression Management mean *T* scores for the PAISE were similar to those of the PAI's normative samples. They also identified strong convergence between Negative and Positive Impression Management scales across the two measures and found that the scales were robust in detecting feigning. However, the authors did not consider clinical, treatment consideration, or interpersonal scales. Regarding concurrent validity, Fantoni-Salvador and Rogers (1997) found that the PAISE had an average correct classification hit rate of .72 when distinguishing Alcohol Dependence, Anxiety Disorders, Major Depression, and Schizophrenia among their clinical sample of Latin American, Mexican American, and Puerto Rican Spanish-speakers. Notably, the three Hispanic groups differed on the Anxiety, Schizophrenia, and Alcohol scales, highlighting the importance of examining country of origin. However, after incorporating participants' number of reported symptoms, no differences in scale elevations across the three groups were significant. While this research provides beginning support for the psychometric properties of the PAISE, it does not constitute a comprehensive validation study and the increased rate of invalid protocols found by Rogers et al. (1995) is particularly concerning. Lack of a thorough research base creates difficulty for practitioners in the United States to engage in informed test selection (Fernandez et al., 2007). It is worth exploring other forms of the PAI with Spanish-speaking populations while research on the PAISE continues to build.

Personality Assessment Inventory European-Spanish Edition (PAIE-S)

Ortiz-Tallo et al. (2011) adapted items per the International Test Commission's guidelines when developing the PAIE-S. The authors found that some PAI and PAISE items were irrelevant or inapplicable to European Spanish-speakers and adapted them accordingly. The authors validated the measure with Spanish normative community and clinical samples that were stratified by sex and age and included individuals from across Spain. As reported by Ortiz-Tallo and colleagues, the PAIE-S sample scored higher on Anxiety, Paranoia, and Treatment Rejection scales and lower on the Warmth scale relative to PAI normative data (Morey, 1991–2007), which the authors took into account when standardizing the measure. They also found good test-retest reliability and convergence

between parallel scales on the Minnesota Multiphasic Inventory–2 Restructured Form (Ben-Porath & Tellegen, 2008) and Millon Clinical Multiaxial Inventory–III (Millon, Davis, & Millon, 1994). Internal consistencies ranged from .46 (Anxiety-Affective) to .89 (Anxiety), with Cronbach's alphas generally over .70. Given its generally robust properties among its validation sample, the PAIE-S may be appropriate for use with Latin American Spanish-speakers as well.

However, there is significant heterogeneity across Hispanic subcultures and the PAIE-S may not generalize to all Spanish-speaking groups. Recognizing this, the PAIE-S developers sought data from a South American sample and examined the PAIE-S among Chileans (Ortiz-Tallo, Cardenal, Ferragut, & Santamaría, 2015). Compared to the Spanish normative sample, Chilean participants had significantly higher *T* scores on approximately 72% of scales (13 out of 18 scales, validity scales were not examined) and 61% of subscales (19 out of 31 subscales), most notably Mania and Mania-Grandiosity. Internal consistencies for the Chilean sample ranged from .66 (Stress) to .86 (Somatic Complaints) with an average alpha of .77. Ortiz-Tallo et al. concluded that differences in test scores and internal consistencies across their Spanish and Chilean samples provided evidence that the PAIE-S should be standardized separately for the Chilean population.

Stover, Solano, and Liporace (2015) administered the PAIE-S to Argentinian Spanish-speakers. In their study, Spanish and Argentinian experts in psychopathology and psychometrics evaluated PAIE-S items, assessing item clarity and whether they held the same meanings across Spain and Argentina. They opined that four items were not appropriate for use with Argentinian populations and adapted the PAIS-E phrasing for these items accordingly. Among the primary scales, they found that alphas ranged from .60 (Stress) to .86 (Anxiety) with an average alpha of .76. Hence, the authors found internal consistencies comparable to Ortiz-Tallo et al.'s (2015) Chilean sample. Still, it is important to remember that Stover et al. assessed for appropriateness of items and made minor item modifications, while Ortiz-Tallo et al. directly administered the PAIE-S to their Chilean sample with no adaptation. No other studies have examined the PAIE-S among Spanish-speakers. To date, the PAIE-S has been researched among Spanish, Chilean, and Argentinian populations. Further research is needed to examine whether these findings generalize across Latin American cultures.

The Current Study

Research indicates that the PAI psychometric properties can vary across groups (e.g., Fantoni-Salvador &

Rogers, 1997). Appropriate psychological assessment of Spanish-speakers is critical. Given that the PAIE-S adaptation followed recommended guidelines by the International Test Commission and has shown some promise with South American Spanish-speakers, it may be appropriate with more diverse groups of Latin American Spanish-speakers as well. We sought to assess the internal consistency, protocol scores, and convergent validity of the PAI and PAIE-S among English/Spanish bilinguals of Latin American descent. We compared the PAI and the PAIE-S, as these measures have both been validated among their respective samples. Findings will inform theories of individual (cultural and linguistic) differences and psychodiagnostic assessment.

Method

Participants

The initial group of participants ($N = 142$) were undergraduate students at a Hispanic-serving, urban university in the northeast United States over a two-semester period. We invited bilingual students who were of Latin American descent and spoke Spanish as a first and/or primary language to participate in the study.

Measures

Reading Level Indicator Spanish-Companion (RLIS-C). The RLIS-C (Williams, 2000) is a 40-item multiple-choice reading screen used to estimate an individual's abilities to read and comprehend Spanish. The measure has been used in past PAI research with bilingual samples (Fernandez et al., 2008). Total raw ability scores are used to indicate participants' reading level (second grade and beyond). The RLIS-C has good test-retest reliability ($r = .90$). To increase participation and minimize burden on participants, we did not administer an English proficiency exam given that participants were attending a four-year English-speaking college and were likely proficient in English.

Demographic Questionnaire. Participants completed a questionnaire (English language format) that queried demographic information such as age, sex, race, ethnicity, and place of birth. They also reported their linguistic experience: first language, age of English fluency, and primary language spoken at home.

Personality Assessment Inventory (PAI) and European-Spanish version (PAIE-S). The PAI and the PAIE-S are both self-report measures of personality and psychopathology that consist of 344 items. Individuals indicate whether items pertain to their personal experience (*false, slightly true, mainly true, or very true*). It should be noted that, while almost all scales and subscales have

the same number of items on both measures, there is one scale in which the number of items differ. The PAI Inconsistency scale has only 20 items, comprised of 10 contradictory item pairs. The PAIE-S developers found that the Inconsistency scale had better discriminant validity when they drew an additional 10 contradictory item pairs from the larger item pool. Consequently, the PAIE-S Inconsistency scale has 40 items. Scales and subscales are reported as T scores ($M = 50$ with $SD = 10$) and are considered to be at clinical elevations when $T \geq 70$. The PAI has good overall internal consistency and reliability (mean alpha = .70 and test-retest correlations = .85). The PAIE-S has demonstrated strong psychometric properties with its normative sample (Ortiz-Tallo et al., 2011). Morey's (2003) recommended validity cut scores for the PAI are Inconsistency $\geq 73\ T$, Infrequency $\geq 75\ T$, Negative Impression Management $\geq 92\ T$, and Positive Impression Management $\geq 68\ T$. For the PAIE-S, protocols are considered invalid if Inconsistency $\geq 75\ T$, Infrequency $\geq 75\ T$, Negative Impression Management $\geq 101\ T$, or Positive Impression Management $\geq 65\ T$ (Ortiz-Tallo et al., 2011).

Procedure

Research assistants, one of whom spoke Spanish, screened the initial 142 participants using the RLIS-C and excluded 12 participants who did not pass the measure at the fourth-grade level. We counterbalanced PAI and PAIE-S administration across two sessions spaced approximately two weeks apart. Participants received course credit upon study completion. Participants' PAI and PAIE-S responses were entered into their respective scoring programs and exported as individual item scores and scale and subscale T scores.

Results

Participants were approximately 20-years-old ($M = 19.83$, $SD = 2.61$). Most ($n = 108$, 85.0%) were female and all participants identified as Hispanic. Over half were born in the United States ($n = 71$, 54.6%), followed by the Dominican Republic ($n = 24$, 18.9%), Ecuador ($n = 11$, 8.7%), Mexico ($n = 7$, 5.5%), and Colombia ($n = 5$, 3.9%), with the remaining participants ($n = 9$, 7.1%) originating from other Latin American countries. Participants who were born outside of the United States reported moving to the United States at 11.27 years old ($SD = 5.37$). Most participants ($n = 69$, 53.1%) said that they spoke Spanish in their household. During the second semester of data collection, we asked 88 participants about their preferred language. The majority cited English as their preferred language ($n = 47$, 56.0%), followed by Spanish ($n = 22$, 26.2%), and then both English and Spanish

(n = 15, 17.9%). Spanish reading level was beyond recommended proficiency (RLIS-C M = 31.42, SD = 3.67).

Validity Analyses

One hundred and ten participants (84.6%) completed the PAI, 106 participants (89.2%) completed the PAIE-S, and 86 (66.2%) completed both measures. In looking at validity scales (Inconsistency, Infrequency, Negative Impression Management, and Positive Impression Management), we found 90 (81.8%) valid PAI protocols and 77 (72.6%) valid PAIE-S protocols. The Inconsistency scale was the most frequently invalid scale for the PAI (9.0% of all PAI protocols), while Infrequency was the most frequently invalid scale for the PAIE-S (18.6% of all PAIE-S protocols). Of the 86 who completed both the PAI and PAIE-S, 76 (88.4%) were both valid for Inconsistency, 66 (76.7%) were both valid for Inconsistency, 86 (100%) were both valid for Negative Impression Management, and 74 (86.0%) were both valid for Positive Impression Management (see Table 1).

However, only one validity scale beyond the aforementioned cut scores rendered a protocol invalid. Among participants who took both measures, 53 (61.6% of 86) participants produced valid protocols for both measures. We proceeded analyses with only valid protocols. We conducted t-tests to assess whether there were significant differences associated with order of PAI/PAIE-S administration. Given the number of comparisons, we used a Bonferroni correction of $p < .00094$ ($p = .05/53$ planned comparisons) to detect significance. We were missing counterbalancing data for five participants. Among the remaining n = 48, there were no significant differences on scale or subscale scores related to PAI/PAIE-S administration order.

Internal Consistency

For the PAI, 16 (72.72%) scales and 11 (35.48%) subscales had alphas above .70 (see Table 2). This indicates that among the valid protocols in our sample, approximately 27.28% of PAI scales and 64.52% PAI subscales

did not have adequate internal consistency. PAI alpha coefficients ranged from .10 (Infrequency) to .89 (Aggression), with an average alpha of .65. With respect to PAIE-S protocols, 11 (50.00%) scales and 8 (25.81%) subscales were at or above .70 alpha, with an average alpha of .65. This indicates that 50.00% of PAIE-S scales and 74.19% of PAIE-S subscales did not have adequate internal consistency. PAIE-S scale and subscales alphas ranged from .27 (Mania-Activity Level) to .81 (Somatic) with an average alpha of .63.

On average, mean PAIE-S inter-item correlations were less robust, and more variable, compared to PAI mean inter-item correlations (Table 2). PAI scale and subscale mean inter-item correlations ranged between r = .02 (Infrequency) and r = .53 (Aggression-Aggressive Attitude), with an overall average of r = .21. PAIE-S scale and subscale inter-item correlations ranged from r = .05 (Mania-Activity Level) to r = .36 (Anxiety-Related Disorders-Traumatic Stress), with an overall average of r = .18.

Scale and Subscale T-scores

Table 3 depicts the range of T scores for scales and subscales for the 53 valid protocols.

PAI mean scale and subscale T scores ranged from a low of 47.19 (SD = 7.45, Alcohol Problems) to a high of 61.15 (SD = 10.56; Paranoia-Hypervigilance). Participants' average PAI T score across scales and subscales was approximately 50 (M = 52.13, SD = 9.12). PAIE-S average T scores ranged from 45.49 (SD = 10.55; Warmth) to 57.64 (SD = 8.92; Paranoia). Similar to the PAI, the mean scale and subscale T score for PAIE-S protocols was approximately 50 (M = 51.47, SD = 9.05). We again used a Bonferroni correction ($p < .00094$) to identify whether or not participants' scale and subscale scores were significantly different across the two measures. We noted 10 instances. On average, participants scored significantly higher on PAI, compared to PAIE-S, on the following: Anxiety, Anxiety-Cognitive, Anxiety-Physiological, Anxiety-related Disorders-Traumatic Stress, Depression-Cognitive, Paranoia-Hypervigilance, Borderline Features, Borderline Features-Affective

Table 1. *Frequencies of Valid PAI and PAIE-S Protocols N = 86*

	PAIE-S INC Valid n = 81 (94.2%)	PAIE-S INF Valid n = 70 (81.4%)	PAIE-S NIM Valid n = 86 (100%)	PAIE-S PIM Valid n = 76 (88.4%)
PAI INC Valid n = 80 (93.0%)	76 (88.4%)	65 (75.6%)	80 (93.0%)	71 (82.6%)
PAI INF Valid n = 82 (95.3%)	77 (89.5%)	66 (76.7%)	82 (95.3%)	71 (82.6%)
PAI NIM Valid n = 86 (100%)	81 (94.2%)	70 (81.4%)	86 (100%)	76 (88.4%)
PAI PIM Valid n = 81 (94.2%)	76 (88.4%)	66 (76.7%)	81 (94.2%)	74 (86.0%)

Note. PAI validity cut scores = ICN ≥ 73t; INF ≥ 75t; NIM ≥ 92t; and PIM ≥ 68t.
PAIE-S validity cut scores = ICN ≥ 75t; INF ≥ 75t; NIM ≥ 101t; and PIM ≥ 65t.

Table 2. *Scale and Subscale Alpha Coefficients and Inter-item Correlations for Valid PAI and PAIE-S Protocols (N = 53)*

Scale/subscale (*n* items)	PAI α	PAIE-S α	PAI *r*	PAIE-S *r*
Validity				
ICN (20/40)[†]	.46	.66	.07	.06
INF (8)	.10	.35	.02	.09
NIM (9)	.51	.50	.13	.11
PIM (9)	.68	.71	.19	.22
Clinical				
SOM (24)	.80	.81	.19	.19
SOM-C (8)	.69	.67	.26	.25
SOM-S (8)	.48	.59	.13	.18
SOM-H (8)	.70	.62	.32	.26
ANX (24)	.86	.87	.22	.23
ANX-C (8)	.73	.72	.24	.25
ANX-A (8)	.68	.70	.23	.24
ANX-P (8)	.60	.67	.17	.22
ARD (24)	.71	.71	.11	.12
ARD-O (8)	.65	.67	.20	.21
ARD-P (8)	.46	.53	.09	.13
ARD-T (8)	.69	.77	.27	.36
DEP (24)	.82	.79	.19	.16
DEP-C (8)	.75	.60	.29	.18
DEP-A (8)	.78	.66	.33	.27
DEP-P (8)	.53	.62	.12	.16
MAN (24)	.76	.64	.12	.07
MAN-A (8)	.23	.27	.04	.05
MAN-G (8)	.66	.51	.21	.12
MAN-I (8)	.81	.76	.34	.28
PAR (24)	.71	.75	.10	.11
PAR-H (8)	.69	.68	.22	.19
PAR-P (8)	.59	.48	.16	.11
PAR-R (8)	.56	.57	.15	.15
SCZ (24)	.79	.80	.14	.16
SCZ-P (8)	.42	.58	.11	.19
SCZ-S (8)	.78	.76	.30	.28
SCZ-T (8)	.79	.76	.34	.30
BOR (24)	.84	.80	.19	.15
BOR-A (6)	.74	.74	.33	.33
BOR-I (6)	.63	.60	.22	.22
BOR-N (6)	.56	.42	.19	.11
BOR-S (6)	.43	.50	.13	.13
ANT (24)	.79	.59	.14	.06
ANT-A (8)	.56	.36	.14	.07
ANT-E (8)	.55	.49	.14	.10
ANT-S (8)	.70	.37	.24	.09
ALC (12)	.72	.72	.28	.21
DRG (12)	.42	.68	.12	.25
Treatment Consideration				
AGG (18)	.89	.80	.33	.22
AGG-A (6)	.87	.70	.53	.30
AGG-V (6)	.63	.64	.22	.30
AGG-P (6)	.73	.48	.42	.14
SUI (12)	.77	.55	.41	.18

Table 2. *(Continued)*

Scale/subscale (*n* items)	PAI α	PAIE-S α	PAI *r*	PAIE-S *r*
STR (8)	.31	.57	.07	.16
NON (8)	.79	.63	.32	.18
RXR (8)	.76	.66	.28	.21
Interpersonal				
DOM (12)	.79	.69	.23	.14
WRM (12)	.76	.72	.21	.18
Average	.68	.65	.21	.18

Note. [†] PAI ICN = 20 items. PAIE-S = 40 items. For PAI clinical scales (bolded), SOM = Somatic Complaints (SOM-C = Conversion, SOM-S = Somatization, and SOM-H = Health Concerns); ANX = Anxiety (ANX-C = Cognitive, ANX-A = Affective, and ANX-P = Physiological); ARD = Anxiety-Related Disorders (ARD-O = Obsessive-Compulsive, ARD-P = Phobia, and ARD-T = Traumatic Stress); DEP = Depression (DEP-C = Cognitive, DEP-A = Affective, and DEP-P = Physiological); MAN = Mania (MAN-A = Activity Level, MAN-G = Grandiosity, and MAN-I = Irritability); PAR = Paranoia (PAR-H = Hypervigilance, PAR-P = Persecution, and PAR-R = Resentment); SCZ = Schizophrenia (SCZ-P = Psychotic Experiences, SCZ-S = Social Detachment, and SCZ-T = Thought Disorder); BOR = Borderline Features (BOR-A = Affective Instability, BOR-I = Identity Problems, BOR-N = Negative Relationships, and BOR-S = Self- Harm); ANT = Antisocial Features (ANT-A = Antisocial Behaviors, ANT-E = Egocentricity, and ANT-S = Stimulus Seeking); ALC = Alcohol Problems, and DRG = Drug Problems. For PAI treatment consideration scales (bolded), AGG = Aggression (AGG-A = Aggressive Attitude, AGG-V = Verbal Aggression, and AGG-P = Physical Aggression); SUI = Suicidal Ideation; STR = Stress; NON = Nonsupport; and RXR = Treatment Rejection. For PAI interpersonal scales (bolded), DOM = Dominance and WRM = Warmth. For PAI validity scales; ICN = Inconsistency; INF = Infrequency; PIM = Positive Impression; and NIM = Negative Impression.

Instability, and Borderline Features-Negative Relationships. They scored significantly lower on the PAI Aggression-Aggressive Attitude subscale compared to the PAIE-S.

To explore why *T* scores varied across the two measures, we conducted post hoc analyses of PAI and PAIE-S raw score item endorsements using $p < .00094$ to detect statistical significance. We did not compare findings across Inconsistency scores given that the PAIE-S has twice as many Inconsistency items than the PAI. Results showed that mean *T* scores were significantly higher on the PAI relative to the PAIE-S but that mean raw scores were *not* significantly different for the Anxiety and Borderline Features scales and the Anxiety-Cognitive, Anxiety-Physiological, Anxiety-Related Disorders-Traumatic Stress, Depression-Cognitive,

Paranoia-Hypervigilance, and Borderline Features-Affective Instability subscales. Mean Aggression-Aggressive Attitude *T* scores were significantly higher on the PAIE-S than the PAI, but there was no significant difference across Aggression-Aggressive Attitude raw scores. Borderline Features-Negative Relationship subscale mean *T* and raw scores were both significantly higher on the PAI relative to the PAIE-S. The Warmth scale was the only scale that showed significant differences between mean raw scores, but not mean *T* scores, with participants typically endorsing more items on the PAI relative to the PAIE-S.

Convergence

Table 3 also shows the convergence statistics for participants who completed valid protocols for both measures (*n* = 53). Nearly all scales converged at *p* < .00094. The Antisocial-Egocentricity subscales converged at *p* = .0011. Inconsistency and Infrequency scales had the lowest convergence statistics, with *r* = .19 and *r* = .20, respectively. Neither relationship achieved statistical significance.

Discussion

Cultural and linguistic diversity is of growing importance in the assessment field. In particular, there is a pressing need to study measures among Latin Americans and Spanish-speakers given the anticipated population increase and professional assessment guidelines. We explored the internal consistency and convergent validity of the PAI and PAIE-S with Latin American Spanish-speakers. Our study is in the spirit of test development and multicultural guidelines by the APA (American Psychological Association, 2017) and Standards for Educational and Psychological Testing (American Education Research Association, American Psychological Association, & National Counsel on Measurement in Education, 2014) and builds on the limited research in this area.

Similar to Rogers' et al. (1995) study of the PAISE with Mexican Americans, relative to Morey (1991–2007) we found an unusually high number of invalid protocols. Participants who produced invalid protocols most often scored above recommended validity cut scores on the PAI Inconsistency or PAIE-S Infrequency scales. We recognize that participants may not have fully attended to the measure, but also consider linguistic issues as a factor. Morey (2003) noted that reading difficulties and confusion on the PAI are among the primary reasons for obtaining high Inconsistency and Infrequency scores. The interaction between scale content and cultural factors (e.g., differences in Latin American and Spanish samples) may also play a role. When looking across the measures for all participants

who took both measures, Negative Impression Management had perfect agreement on scale validity. Inconsistency and Positive Impression Management scales also had good agreement, with Infrequency having the least.

With respect to internal consistency, approximately 72% of PAI scales and 35% of PAI subscales met our cutoff of .70 for acceptable internal consistency. Findings were not as encouraging for PAIE-S protocols: 50% of scales and 26% subscales had alphas above .70. In general, mean inter-item correlations were largely similar to Morey's (1991–2007) census sample and the PAIE-S normative sample. However, internal consistencies are significantly lower than PAI and PAIE-S validation samples and suggests that a significant number of PAI and PAIE-S scales and subscales may not adequately measure unitary constructs and bring into question their appropriateness for use with Latin American bilinguals.

As validity is the first step in protocol interpretation, we paid special attention to the measures' validity scales. We noted some differences in alphas across PAI and PAIE-S Inconsistency scales. The PAI's relatively lower alpha could be partly attributed to the number of each scales' test items, given that the PAI consists of 10 inconsistent pairs while the PAIE-S consists of 20 and internal consistency increases with the number of test items. The PAI Infrequency scale's alpha was the lowest of all scale and subscale alpha values, and the PAIE-S alpha was relatively low as well. These findings could indicate that Inconsistency and Infrequency item endorsements were not as unusual among our sample as developers intended and do not capture particularly unusual item endorsement. These differences may be attributable to the sample's unique characteristics relative to those of other research samples. Alternatively, and perhaps more likely, it may simply be that Inconsistency and Infrequency scales do not capture theoretically meaningful constructs and by definition would have low internal consistency (Morey, 1991–2007). Negative Impression Management had relatively low alphas. It may be that our samples' expressions of negative self-evaluation and distress differ from PAI and PAIE-S normative samples. If this is the case, it is unclear why its intended *opposite* scale, Positive Impression Management, had relatively strong and consistent alphas.

Although internal consistency was not ideal for a number of scales, several clinical scales stood out as particularly problematic with α ≤ .40. Mania-Activity Level for both measures was below this cut off, while Antisocial Features-Antisocial Behaviors for the PAIE-S and the Stress scale for PAI was also below α ≤ .40. We suspect that there is a common theme among these

Table 3. *Scale and Subscale T-Scores, Bivariate Comparisons, and Scale Convergence for Valid PAI and PAIE-S Protocols (N = 53)*

Scale/subscale (n items)	PAI M (SD)	PAIE-S M (SD)	T Difference t	p	d	T Convergence r
Validity						
ICN (20/40)[†]	50.30 (7.68)	51.43 (9.23)	−0.76	.44917	−0.10	.19
INF (8)	52.72 (7.42)	51.53 (10.57)	0.74	.46115	0.10	.20
NIM (9)	49.53 (7.65)	51.72 (9.98)	−2.03	.04790	−0.28	.63*
PIM (9)	49.62 (9.19)	50.23 (9.71)	−0.54	.59433	−0.07	.63*
Clinical						
SOM (24)	50.09 (6.99)	49.93 (8.08)	0.22	.82865	0.03	.72*
SOM-C (8)	48.87 (7.89)	50.57 (8.96)	−1.84	.07088	−0.25	.69*
SOM–S (8)	51.06 (7.02)	47.93 (8.28)	3.32	.00165	0.46	.61*
SOM-H (8)	49.96 (7.82)	51.26 (7.99)	−1.72	.09120	−0.24	.76*
ANX (24)	54.23 (8.59)	49.36 (8.23)	6.38	< .00001	0.88*	.78*
ANX-C (8)	54.53 (8.66)	49.51 (8.78)	6.17	< .00001	0.85*	.77*
ANX-A (8)	52.77 (9.02)	50.32 (8.20)	2.47	.01678	0.34	.65*
ANX-P (8)	54.19 (9.06)	48.51 (8.18)	5.362	< .00001	0.74*	.60*
ARD (24)	51.75 (8.90)	51.15 (7.37)	0.70	.48729	0.10	.72*
ARD-O (8)	50.32 (10.32)	52.26 (10.08)	−2.26	.02814	−0.31	.81*
ARD-P (8)	53.53 (9.29)	53.87 (8.93)	−0.34	.73814	−0.05	.68*
ARD-T (8)	50.19 (7.84)	46.89 (7.27)	3.62	.00066	0.50*	.62*
DEP (24)	51.96 (8.82)	51.13 (8.18)	1.03	.31031	0.14	.76*
DEP-C (8)	54.32 (10.58)	49.66 (9.51)	4.70	.00002	0.65*	.75*
DEP-A (8)	51.45 (9.85)	50.26 (8.56)	1.38	.17507	0.19	.78*
DEP-P (8)	49.57 (8.20)	52.70 (8.69)	−2.91	.00532	−0.40	.57*
MAN (24)	53.75 (9.52)	53.4 0(7.27)	0.37	.71262	0.05	.68*
MAN-A (8)	51.06 (8.54)	52.00 (7.12)	−0.82	.41462	−0.11	.44*
MAN-G (8)	54.85 (10.08)	56.23 (8.50)	−1.29	.20266	−0.18	.66*
MAN-I (8)	52.57 (10.89)	49.43 (9.50)	2.82	.00674	0.39	.69*
PAR (24)	59.40 (7.78)	57.64 (8.92)	2.34	.02316	0.32	.79*
PAR-H (8)	61.15 (10.56)	56.89 (9.26)	3.97	.00022	0.55*	.70*
PAR-P (8)	56.98 (7.57)	56.94 (9.69)	0.03	.97425	0.004	.54*
PAR-R (8)	55.62 (9.46)	55.32 (9.46)	0.23	.81731	0.03	.50*
SCZ (24)	51.62 (9.93)	52.85 (9.94)	−1.07	.28934	−0.15	.65*
SCZ-P (8)	48.58 (8.01)	51.72 (9.12)	−2.81	.00689	−0.39	.56*
SCZ-S (8)	52.96 (10.81)	54.26 (10.96)	−1.03	.30659	−0.14	.65*
SCZ-T (8)	51.45 (11.85)	50.09 (9.99)	1.11	.27062	0.15	.68*
BOR (24)	54.85 (9.58)	51.45 (8.92)	3.59	.00073	0.49*	.73*
BOR-A (6)	52.96 (10.61)	48.91 (8.90)	3.71	.00051	0.51*	.68*
BOR-I (6)	53.83 (9.58)	52.25 (9.89)	1.39	.16997	0.19	.64*
BOR-N (6)	56.11 (9.73)	50.89 (9.17)	4.08	.00016	0.56*	.51*
BOR-S (6)	52.26 (9.06)	54.06 (9.35)	−1.65	.10603	−0.23	.63*
ANT (24)	52.43 (8.99)	52.87 (8.18)	−0.37	.71431	−0.05	.50*
ANT-A (8)	47.62 (7.55)	50.91 (8.14)	−2.90	.00553	−0.40	.45*
ANT-E (8)	55.74 (9.76)	54.94 (10.97)	0.52	.60403	0.07	.44
ANT-S (8)	54.03 (11.01)	51.87 (8.48)	2.05	.04542	0.28	.63*
ALC (12)	47.19 (7.45)	49.94 (10.40)	−3.07	.00342	−0.42	.78*
DRG (12)	47.92 (6.36)	47.49 (8.90)	0.41	.68529	0.06	.53*
Treatment Consideration						
AGG (18)	50.64 (11.43)	50.74 (9.78)	−0.10	.92069	−0.01	.80*
AGG-A (6)	48.91 (12.70)	53.57 (9.98)	−4.32	.00007	−0.59*	.79*
AGG-V (6)	52.3 0(9.94)	49.43 (9.73)	2.26	.02793	0.31	.56*
AGG-P (6)	50.25 (10.25)	48.94 (9.35)	1.48	.14478	0.20	.79*
SUI (12)	49.26 (7.86)	50.06 (7.81)	−1.26	.21506	−0.17	.83*

Table 3. *(Continued)*

| Scale/subscale (n items) | PAI | PAIE-S | T Difference | | | T Convergence |
	M (SD)	M (SD)	t	p	d	r
STR (8)	51.17 (6.36)	52.51 (8.46)	–1.23	.22512	–0.17	.46*
NON (8)	52.15 (11.05)	52.83 (9.83)	–0.59	.55606	–0.08	.69*
RXR (8)	51.53 (9.40)	49.09 (8.91)	2.36	.02184	0.32	.67*
Interpersonal						
DOM (12)	52.83 (9.10)	45.49 (10.55)	0.26	.79911	0.04	.72*
WRM (12)	47.60 (9.69)	52.59 (9.63)	2.28	.02649	0.31	.78*
Average	52.97(9.92)	52.25(9.52)	0.72	.27708	0.09	.64

Note. [†] PAI ICN = 20 items. PAIE-S = 40 items. Convergence = correlation between the Spanish and English PAIs. For PAI clinical scales (bolded), SOM = Somatic Complaints (SOM-C = Conversion, SOM-S = Somatization, and SOM-H = Health Concerns); ANX = Anxiety (ANX-C = Cognitive, ANX-A = Affective, and ANX-P = Physiological); ARD = Anxiety- Related Disorders (ARD-O = Obsessive-Compulsive, ARD-P = Phobia, and ARD-T = Traumatic Stress); DEP = Depression (DEP-C = Cognitive, DEP-A = Affective, and DEP-P = Physiological); MAN = Mania (MAN-A = Activity Level, MAN-G = Grandiosity, and MAN-I = Irritability); PAR = Paranoia (PAR-H = Hypervigilance, PAR-P = Persecution, and PAR-R = Resentment); SCZ = Schizophrenia (SCZ-P = Psychotic Experiences, SCZ-S = Social Detachment, and SCZ-T = Thought Disorder); BOR = Borderline Features (BOR-A = Affective Instability, BOR-I = Identity Problems, BOR-N = Negative Relationships, and BOR-S = Self- Harm); ANT = Antisocial Features (ANT-A = Antisocial Behaviors, ANT-E = Egocentricity, and ANT-S = Stimulus Seeking); ALC = Alcohol Problems, and DRG = Drug Problems. For PAI treatment consideration scales (bolded), AGG = Aggression (AGG-A = Aggressive Attitude, AGG-V = Verbal Aggression, and AGG-P = Physical Aggression); SUI = Suicidal Ideation; STR = Stress; NON = Nonsupport; and RXR = Treatment Rejection. For PAI interpersonal scales (bolded), DOM = Dominance and WRM = Warmth. For PAI validity scales, ICN = Inconsistency; INF = Infrequency; PIM = Positive Impression; and NIM = Negative Impression.
*$p < .00094$.

scales – active and risk-taking behaviors of early adulthood (mean age for our sample), which along with stress of college and/or acculturation could interfere with these scales' intended constructs. It is unclear why internal consistency would be low in PAIE-S Antisocial Features-Antisocial Behavior but not on the PAI, and why the PAI Stress scale would be so much lower than on the PAIE-S.

Some of our findings that differ from past research may be explained by differences in sampling and methodology. We found lower internal consistency on average for PAI scales than Rogers et al. (1995); however, their sample originated from Mexico, while our sample was more diverse, with only 5.50% of our sample identifying as of Mexican descent. The notable heterogeneity among Hispanic cultures (Sue & Sue, 2012) may have been reflected in our sample's item endorsement relative to Rogers et al.'s sample of Mexican participants. Unfortunately, sample size limited exploring PAI and PAIE-S scores as a function of country of origin in our study. Other possible explanations for differences in the PAI's internal consistency compared to Rogers et al. (1995) may be the age (younger), sex (largely female), education (college attendance), and context (college versus clinical setting) of our sample, as well as changing societal norms over the course of 20 years. Our sample was much younger than samples

in other PAIE-S studies (Ortiz-Tallo et al., 2015; Stover et al., 2015).

Despite relatively low internal consistency, *T*-scores for Infrequency, Mania-Activity Level, Antisocial Features-Antisocial Behavior, and Stress did not significantly differ, evidencing that there is some overlap in these constructs across measures. Findings for mean PAI Positive and Negative Impression Management *T* scores were similar to those of Fernandez et al. (2008), in that they were similar to the PAI's normative sample. Also in line with past research, participants in this study tended to score consistently low *T* scores on the Warmth scale and high *T* scores on Paranoia subscales. These findings closely mirror those of others (Ortiz-Tallo et al., 2015; Ortiz-Tallo et al., 2011) and suggest that Latin American, South American, and European-Spanish subgroups may report relatively low Warmth characteristics and relatively high, but not necessarily clinically elevated, symptoms and experiences associated with paranoia. Indeed, some researchers have reported relatively elevated paranoia scores in non-U.S. samples (Groves & Engel, 2007).

There still were a number of scales and subscales in which we observed significantly different profile scores. Specifically, we found that on average, participants obtained significantly higher *T* scores on PAI Anxiety, Anxiety-Cognitive, Anxiety-Physiological,

Table 4. *Scale and Subscale Raw Scores and Bivariate Comparisons for Valid PAI and PAIE-S Protocols (N = 53)*

Scale/subscale (n items)	PAI M (SD)	PAIE-S M (SD)	Raw Difference t	p	d
Validity					
ICN (20/40)[†]	5.43 (2.56)	11.36 (3.98)	−10.06	< .00001	−1.38*
INF (8)	3.38 (1.92)	3.42 (2.37)	−0.10	.92127	−0.10
NIM (9)	1.58 (2.06)	2.08 (2.33)	−1.89	.06404	−0.26
PIM (9)	14.92 (3.99)	15.87 (4.24)	−1.93	.05876	−0.27
Clinical					
SOM (24)	11.09 (6.99)	12.17 (7.81)	−1.43	.15828	−0.20
SOM–C (8)	2.06 (2.76)	2.87 (3.03)	−2.57	.01299	−0.35
SOM-S (8)	4.92 (2.67)	4.30 (3.30)	1.68	.09891	0.23
SOM-H (8)	4.11 (3.30)	5.00 (3.11)	−2.88	.00576	−0.40
ANX (24)	20.92 (9.10)	20.11 (9.22)	0.98	.33329	0.13
ANX-C (8)	8.00 (3.76)	7.32 (3.58)	1.98	.05273	0.27
ANX-A (8)	7.28 (3.47)	7.89 (3.76)	−1.46	.15088	−0.20
ANX-P (8)	5.64 (3.25)	4.91 (3.19)	1.88	.06636	0.26
ARD (24)	21.38 (7.40)	22.55 (7.76)	−1.48	.14418	−0.20
ARD-O (8)	9.47 (3.88)	9.55 (4.19)	−0.22	.82761	−0.03
ARD-P (8)	7.96 (3.37)	9.06 (3.58)	−2.83	.00663	−0.39
ARD-T (8)	3.94 (3.25)	3.94 (3.63)	0.00	1.00000	0.00
DEP (24)	16.11 (8.29)	16.62 (7.68)	−0.66	.51001	−0.09
DEP-C (8)	5.81 (3.60)	5.55 (3.15)	0.79	.43472	0.11
DEP-A (8)	4.58 (3.58)	4.38 (3.05)	0.66	.51231	0.09
DEP-P (8)	5.72 (3.43)	6.70 (3.75)	−2.13	.03827	−0.29
MAN (24)	26.53 (8.72)	24.62 (6.48)	2.15	.03587	0.30
MAN-A (8)	7.04 (2.76)	5.98 (2.41)	2.79	.00732	0.38
MAN-G (8)	10.51 (4.40)	10.79 (3.51)	−0.62	.54023	−0.08
MAN-I (8)	8.98 (4.64)	7.85 (4.14)	2.38	.02087	0.33
PAR (24)	26.60 (6.78)	28.13 (8.09)	−2.22	.03087	−0.30
PAR-H (8)	11.42 (3.58)	12.66 (3.68)	−3.21	.00225	−0.44
PAR-P (8)	6.00 (2.54)	4.77 (2.81)	3.50	.00098	0.48
PAR-R (8)	9.19 (3.32)	10.70 (3.70)	−3.12	.00295	−0.43
SCZ (24)	15.25 (7.76)	17.21 (8.07)	−2.14	.03725	−0.29
SCZ-P (8)	3.68 (2.38)	4.11 (2.81)	−1.26	.21336	−0.17
SCZ-S (8)	6.74 (4.26)	7.45 (4.32)	−1.45	.15334	−0.20
SCZ-T (8)	4.83 (3.98)	5.64 (3.79)	−1.89	.06390	−0.26
BOR (24)	22.85 (9.58)	21.43 (8.84)	1.50	.13903	0.21
BOR-A (6)	5.64 (3.50)	5.55 (3.26)	0.25	.80260	0.03
BOR-I (6)	6.26 (3.22)	6.42 (3.30)	−0.40	.69386	−0.05
BOR-N (6)	7.06 (3.08)	5.40 (2.94)	4.07	.00016	0.56*
BOR-S (6)	3.89 (2.35)	4.08 (2.42)	−0.67	.50539	−0.09
ANT (24)	15.4 (8.18)	16.34 (6.36)	−0.92	.35991	−0.13
ANT-A (8)	3.94 (3.37)	5.06 (3.25)	−2.33	.02375	−0.32
ANT-E (8)	5.13 (2.93)	4.87 (2.84)	0.63	.53137	0.09
ANT-S (8)	6.32 (4.06)	6.42 (2.96)	−0.22	.82835	−0.03
ALC (12)	3.32 (4.14)	3.81 (4.33)	−1.28	.20757	−0.18
DRG (12)	2.96 (3.18)	3.09 (4.12)	−0.26	.79316	−0.04
Treatment Consideration					
AGG (18)	15.32 (9.67)	14.59 (7.66)	0.93	.35679	0.13
AGG-A (6)	5.40 (4.57)	6.21 (3.43)	−2.09	.04172	−0.29
AGG-V (6)	7.53 (3.49)	6.53 (3.18)	2.33	.02351	0.32
AGG-P (6)	2.40 (3.01)	1.85 (2.48)	2.12	.03856	0.29
SUI (12)	3.04 (3.78)	2.43 (3.20)	2.12	.03927	0.29

Table 4. *(Continued)*

Scale/subscale (*n* items)	PAI	PAIE-S	Raw Difference		
	M (*SD*)	*M* (*SD*)	*t*	*p*	*d*
STR (8)	6.38 (2.83)	7.32 (3.38)	–2.09	.04175	–0.29
NON (8)	5.68 (4.04)	6.55 (3.44)	–2.11	.04005	–0.29
RXR (8)	14.45 (4.38)	14.79 (4.02)	–0.72	.47283	–0.10
Interpersonal					
DOM (12)	22.15 (5.13)	21.85 (4.73)	0.60	.54982	0.08
WRM (12)	22.13 (5.45)	19.42 (5.57)	5.44	< .00001	0.75*
Average	9.21 (4.37)	9.33 (4.20)	–0.37	.24520	–0.05

Note. † PAI ICN = 20 items. PAIE-S = 40 items. For PAI clinical scales (bolded), SOM = Somatic Complaints (SOM-C = Conversion, SOM-S = Somatization, and SOM-H = Health Concerns); ANX = Anxiety (ANX-C = Cognitive, ANX-A = Affective, and ANX-P = Physiological); ARD = Anxiety-Related Disorders (ARD-O = Obsessive-Compulsive, ARD-P = Phobia, and ARD-T = Traumatic Stress); DEP = Depression (DEP-C = Cognitive, DEP-A = Affective, and DEP-P = Physiological); MAN = Mania (MAN-A = Activity Level, MAN-G = Grandiosity, and MAN-I = Irritability); PAR = Paranoia (PAR-H = Hypervigilance, PAR-P = Persecution, and PAR-R = Resentment); SCZ = Schizophrenia (SCZ-P = Psychotic Experiences, SCZ-S = Social Detachment, and SCZ-T = Thought Disorder); BOR = Borderline Features (BOR-A = Affective Instability, BOR-I = Identity Problems, BOR-N = Negative Relationships, and BOR-S = Self- Harm); ANT = Antisocial Features (ANT-A = Antisocial Behaviors, ANT-E = Egocentricity, and ANT-S = Stimulus Seeking); ALC = Alcohol Problems, and DRG = Drug Problems. For PAI treatment consideration scales (bolded), AGG = Aggression (AGG-A = Aggressive Attitude, AGG-V = Verbal Aggression, and AGG-P = Physical Aggression); SUI = Suicidal Ideation; STR = Stress; NON = Nonsupport; and RXR = Treatment Rejection. For PAI interpersonal scales (bolded), DOM = Dominance and WRM = Warmth. For PAI validity scales, ICN = Inconsistency; INF = Infrequency; PIM = Positive Impression; and NIM = Negative Impression.
*p < .00094.

Anxiety-related Disorders-Traumatic Stress, Depression-Cognitive, Paranoia-Hypervigilance, Borderline Features, Borderline Features-Affective Instability, and Borderline Features-Negative Relationships. They obtained significantly lower *T* scores on the PAI Aggression-Aggressive Attitude subscale compared to the PAIE-S. For each of these significant standardized *T* differences, we did *not* find remarkable discrepancies at the raw score item endorsement level. These findings imply that for each of the significantly different *T* scores on the aforementioned scales and subscales, characteristics and/or experiences were more or less common in either one or the other normative samples. For example, there appears to be a theme of anxiety throughout scales and subscale constructs in which participants scored significantly higher on the PAI. This suggests that anxiety-related themes are more common among our sample than the PAI's normative sample. This may mean that the PAI is either more sensitive to anxiety-related themes in our sample or is overpathologizing our participants. On the other hand, it may also mean that the PAIE-S is underpathologizing our sample or is more reflective of cultural norms. Their higher PAI scores may be attributed to characteristics among Hispanic culture, such as culturally normative anxiety, potentially resulting from the effects of social marginalization, which is more salient on an

English language measure (Gamst et al., 2002; Hiott, Grzywacz, Arcury, & Quandt, 2006; Sue & Sue, 2012). However, research indicates that bilinguals tend to experience emotion more strongly in native rather than foreign languages (Caldwell-Harris, 2015). Then again, in some instances, Latin Americans tend to underreport anxiety and depression (Bell et al., 2011; Leung, LaChapelle, Scinta, & Olvera, 2014). Taking these findings under consideration, it is unclear precisely why our bilingual participants tended to score higher on the PAI's anxious themes. While it may be that the PAI overpathologized our sample in some respects, it may be also that the PAIE-S *underpathologized* the sample.

Virtually all scales (aside from Inconsistency and Infrequency) and subscales converged at our Bonferroni correction with the exception of Antisocial Features-Egocentricity. While Ortiz-Tallo and colleagues (2015) found that Chileans were significantly more likely to endorse items that produced higher PAIE-S Mania scores compared to their normative sample, this was not the case in our study. Instead, our participants tended to endorse PAI items broadly related to anxiety. Taken together, findings support that socially normative thoughts, feelings, and behaviors may vary across Hispanic subcultures, or at least across language format. It could be that these demographic groups differ

in their experience of mental health symptoms, their likelihood of reporting psychiatric distress, or other unknown factors that may influence measurement outcomes. Our findings may also be impacted by sample size and participant attrition. We retained two-thirds of the small pool of eligible participants for our convergence analyses, and only two-thirds of that sample produced valid protocols.

Implications for Practice and Research

Practitioners may use these results to inform their test selection and interpretation. Fernandez et al. (2007) encourage practitioners to consider their testing options, the research regarding these tests, the tests' applicability to clients, and their confidence in the testing process. Our findings suggest that practitioners should take extra care when assuming a test's generalizability across cultural subgroups. Notably, participants in our sample were all enrolled in an English-speaking four-year college. In clinical practice, some practitioners may assume that English-speaking college students are entirely English proficient and that they may appropriately administer and interpret the PAI using its normative scores. We did not screen for English proficiency. It could be that some participants were not as strong in their English reading comprehension as we anticipated, which could have influenced our findings. We encourage practitioners to consider linguistic preferences and proficiencies in their test selection and utilize tools to assess reading level before administering English-language tests to bilinguals. In addition, researchers may wish to explore whether or not high paranoia scale and subscales findings are common among Hispanic individuals residing in the U.S. Perceived prejudice and discrimination could drive higher reported paranoia, as these elevations can indicate suspiciousness, mistrust, hypervigilance, and beliefs that one has been mistreated.

Our study, in collection with others (Ortiz-Tallo et al., 2015; Ortiz-Tallo et al., 2011; Rogers et al., 1995; Stover et al., 2015) suggests that the PAI and PAIE-S sometimes function differently across Spanish-speaking subgroups. Continued research is needed to expand our understanding of personality assessment as it pertains to diverse groups of Spanish-speakers. Future researchers, using larger sample sizes, should examine the factor structures of PAI and PAIE-S scales and subscales among English/Spanish bilinguals both in the community and in clinical contexts. In a small study of community-dwelling Latin American English-Spanish speaking bilinguals, Obando, Pearson, Kois, and Chauhan (2016, August) found PAI and PAIE-S psychometric properties similar to those identified here. Their sample was notably older (M = 50) and

approximately half male and half female. As such, early evidence of generalizability has emerged and clinical replication research is warranted.

We encourage researchers to assess whether country of origin is significantly associated with differences in PAI/PAIE-S protocols. There are several characteristics of our data that, regrettably, limit this approach. Although a country-by-country analysis is ideal, we could not conduct these statistics given the small cell sizes. We would face sample size issues even if we collapsed Central and South American participants into one group and compared them to U.S.-born participants. While 46% of the sample was born outside of the U.S., many of them moved to the U.S. as young children, while many others moved in adolescence or adulthood (M_{age} move to U.S. = 11.27, SD = 5.37, 95% CI [0.53, 22.01 years]). We felt that it was inappropriate to collapse the hypothetical participant who moved to the U.S. at the age of 1 with a participant who moved to the U.S. at age 20.

An item-level analysis, using a larger sample size, will be important to help clarify which items are problematic across groups. This is particularly the case for those scales and subscales in which participants scored significantly different across measures (e.g., on themes of anxiety and aggression). For example, Anxiety and Borderline Features scales both had multiple subscales in which participants scored significantly higher on the PAI relative to the PAIE-S. Discriminant validity studies may help discern whether or not the measures adequately capture Anxiety and Borderline constructs and elucidate potential over/underpathologizing by the PAI and/or PAIE-S.

In addition to validation studies, acculturation is an important factor to consider in future work. Acevedo-Polakovich et al. (2007) remind practitioners to be attentive to clients' immigration histories, acculturation, and acculturative stress. Indeed, researchers have identified associations between personality assessment outcomes and acculturative status across various race and ethnic groups (Chang & Smith, 2015; Tsai & Pike, 2000; Wong, Correa, Robinson, & Lu, 2017). Further research should include acculturative measures and explore whether or not test properties vary according to acculturation status. Regarding study methodology, future researchers can consider whether administration order is associated with participants' PAI/PAIE-S profiles. Lastly, future researchers may consider whether test-retest reliability is an issue, given that both measures tap into similar constructs.

Practitioners should continue to closely consider the psychometric properties of available tests while researchers explore the best means to assess Spanish-speaking populations. We found that participants in our study frequently scored higher on PAI scales

and subscales than the PAIE-S and that PAI and PAIE-S scales and subscales generally exhibited inadequate internal consistency when administered to our sample of Spanish-speaking participants. These findings bring into question whether either measure is entirely appropriate for use with similar samples. It is our hope that these results encourage further research and test development in this area. By continuing to explore testing issues among Latin Americans and Spanish speakers, the practitioner and research communities will do service to increasingly diverse populations who deserve culturally sensitive psychological assessment.

References

Acevedo-Polakovich I. D., Reynaga-Abiko G., Garriott P. O., Derefinko K. J., Wimsatt M. K., Gudonis L. C., & Brown T. L. (2007). Beyond instrument selection: Cultural considerations in the psychological assessment of U.S. Latinas/os. *Professional Psychology: Research and Practice*, *38*(4), 375–384. https://doi.org/10.1037/0735-7028.38.4.375

American Psychological Association (2017). *Multicultural guidelines: An ecological approach to context, identity, and intersectionality*. Washington, DC: Author.

American Education Research Association, American Psychological Association, & National Counsel on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bell R. A., Franks P., Duberstein P. R., Epstein R. M., Feldman M. D., y Garcia E. F., & Kravitz R. L. (2011). Suffering in silence: Reasons for not disclosing depression in primary care. *The Annals of Family Medicine*, *9*(5), 439–446. https://doi.org/10.1370/afm.1277.

Ben-Porath Y. S., & Tellegen A. (2008). *MMPI–2–RF, Minnesota Multiphasic Personality Inventory–2 restructured form: Manual for administration, scoring and interpretation*. Minneapolis, MN: Pearson.

Boscán D. C., Penn N. E., Velásquez R. J., Reimann J., Gómez N., Guzmán M., ... de Romero M. C. (2000). MMPI–2 profiles of Colombian, Mexican, and Venezuelan university students. *Psychological Reports*, *87*(1), 107–110.

Butcher J. N., Dahlstrom W., Graham J., Tellegen A., & Kaemmer B. (1989). *MMPI–2: Manual for administration and scoring*. Minneapolis, MN: University of Minnesota Press.

Caldwell-Harris C. L. (2015). Emotionality differences between a native and foreign language: Implications for everyday life. *Current Directions in Psychological Science*, *24*(3), 214–219. https://doi.org/10.1177/0963721414566268

Chang J., & Smith S. R. (2015). An exploration of how Asian Americans respond on the Personality Assessment Inventory. *Asian American Journal of Psychology*, *6*(1), 25. https://doi.org/10.1037/a0036173

Edens J. F., & Ruiz M. A. (2008). Identification of mental disorders in an in-patient prison psychiatric unit: Examining the criterion-related validity of the Personality Assessment Inventory. *Psychological Services*, *5*(2), 108–117. https://doi.org/10.1037/1541-1559.5.2.108

Estrada A. R., & Smith S. R. (2017). An exploration of Latina/o Respondent Scores on the Personality Assessment Inventory. *Current Psychology*, 1–10. https://doi.org/10.1007/s12144-017-9652-2

Fantoni-Salvador P., & Rogers R. (1997). Spanish versions of the MMPI–2 and PAI: An investigation of concurrent validity with Hispanic patients. *Assessment*, *4*(1), 29–39. https://doi.org/10.1177/107319119700400104

Fernandez K., Boccaccini M. T., & Noland R. M. (2007). Professionally responsible test selection for Spanish-speaking clients: A four-step approach for identifying and selecting translated tests. *Professional Psychology: Research and Practice*, *38*(4), 363–374. https://doi.org/10.1037/0735-7028.38.4.363

Fernandez K., Boccaccini M. T., & Noland R. M. (2008). Detecting over-and underreporting of psychopathology with the Spanish-language Personality Assessment Inventory: Findings from a simulation study with bilingual speakers. *Psychological Assessment*, *20*(2), 189–194. https://doi.org/10.1037/1040-3590.20.2.189

Flores A. (2017). *How the U.S. Hispanic population is changing*. Retrieved from https://www.pewresearch.org/fact-tank/2017/09/18/how-the-u-s-hispanic-population-is-changing/

Gamst G., Dana R. H., Der-Karabetian A., Aragón M., Arellano L. M., & Kramer T. (2002). Effects of Latino acculturation and ethnic identity on mental health outcomes. *Hispanic Journal of Behavioral Sciences*, *24*(4), 479–504. https://doi.org/10.1177/0739986302238216

Gardner B. O., Boccaccini M. T., Bitting B. S., & Edens J. F. (2015). Personality Assessment Inventory scores as predictors of misconduct, recidivism, and violence: A meta-analytic review. *Psychological Assessment*, *27*(2), 534–544. https://doi.org/10.1037/pas0000065

Groves J. A., & Engel R. R. (2007). The German adaptation and standardization of the Personality Assessment Inventory (PAI). *Journal of Personality Assessment*, *88*(1), 49–56. https://doi.org/10.1080/00223890709336834

Gurven M., von Rueden C., Massenkoff M., Kaplan H., & Lero Vie M. (2013). How universal is the Big Five? Testing the five-factor model of personality variation among forager–farmers in the Bolivian Amazon. *Journal of Personality and Social Psychology*, *104*(2), 354–370. https://doi.org/10.1037/a0030841.

Hiott A., Grzywacz J. G., Arcury T. A., & Quandt S. A. (2006). Gender differences in anxiety and depression among immigrant Latinos. *Families, Systems, & Health*, *24*(2), 137–146. https://doi.org/10.1037/1091-7527.24.2.137

Leung P., LaChapelle A. R., Scinta A., & Olvera N. (2014). Factors contributing to depressive symptoms among Mexican Americans and Latinos. *Social Work*, *59*(1), 42–51. https://doi.org/10.1093/sw/swt047

Millon T., Davis R., & Millon C. (1994). *MCMI–III™. Manual. Millon Clinical Multiaxial Inventory–III*. Minneapolis, MN: National Computer Systems.

Morey L. (1992). *Tasa de Inventario de la Personalidad. Versión en español de Carlos J. Cano, asesoramiento técnico de Pedro M. Ferreira* [The Personality Assessment Inventory, Spanish version by Carlos J. Cano, technical

advice by Pedro M. Ferreira]. Lutz, FL: Psychological Assessment Resources.

Morey L. (2003). *Essentials of PAI assessment*. Hoboken, NJ: John Wiley & Sons Inc.

Morey L. C. (1991–2007). *Personality Assessment Inventory* [Measurement instrument]. Odessa, FL: Psychological Assessment Resources.

Obando V., Pearson J., Kois L., & Chauhan P. (2016, August). Personality Assessment Inventory English and European Spanish versions with a community sample of Latin American Spanish speakers. *Paper presented at the American Psychology-Law Society*. Atlanta, GA.

Ortiz-Tallo M., Cardenal V., Ferragut M., & Santamaría P. (2015). Spanish and Chilean standardizations of the Personality Assessment Inventory: The influence of sex. *The Spanish Journal of Psychology*, *18*, e48. https://doi.org/10.1017/sjp.2015.57.

Ortiz-Tallo M., Santamaría P., Cardenal V., & Sánchez M. P. (2011). *Adaptación al español del PAI (Personality Assessment Inventory) de L. Morey* [Adaptation to Spanish of PAI (Personality Assessment Inventory) from L. Morey] (Vols. I–II). Madrid, Spain: TEA Ediciones.

Rogers R., Flores J., Ustad K., & Sewell K. W. (1995). Initial validation of the Personality Assessment Inventory-Spanish version with clients from Mexican American communities. *Journal of Personality Assessment*, *64*(2), 340–348. https://doi.org/10.1207/s15327752jpa6402_12

Simons G. F., & Fennig C. D. (2017). *Ethnologue: Languages of the world (20th Ed.).* Dallas, Texas: SIL International.

Sinclair S. J., Smith M., Chung W.-J., Liebman R., Stein M. B., Antonius D., ... Blais M. A. (2015). Extending the validity of the personality assessment inventory's (PAI) Level of Care Index (LOCI) in multiple psychiatric settings. *Journal of Personality Assessment*, *97*(2), 145–152. https://doi.org/10.1080/00223891.2014.941441

Stover J. B., Solano A. C., & Liporace M. F. (2015). Personality Assessment Inventory: Psychometric Analyses of its Argentinean Version. *Psychological Reports*, *117*(3), 799–823. https://doi.org/10.2466/08.03.PR0.117c27z2

Sue D. W., & Sue D. (2012). *Counseling the culturally diverse: Theory and practice*. Hoboken, NJ: John Wiley & Sons Inc.

Tsai D. C., & Pike P. L. (2000). Effects of acculturation on the MMPI–2 scores of Asian American students. *Journal of Personality Assessment*, *74*(2), 216–230. https://doi.org/10.1207/S15327752JPA7402_4

Weiss R. A., & Rosenfeld B. (2012). Navigating cross-cultural issues in forensic assessment: Recommendations for practice. *Professional Psychology: Research and Practice*, *43*(3), 234–240. https://doi.org/10.1037/a0025850

Williams K. T. (2000). *Reading-Level Indicator Spanish Companion: A quick group Reading Placement Test.* Circle Pines, MN: Pearson.

Wong C. C. Y., Correa A., Robinson K., & Lu Q. (2017). The roles of acculturative stress and social constraints on psychological distress in Hispanic/Latino and Asian immigrant college students. *Cultural Diversity and Ethnic Minority Psychology*, *23*(3), 398–406. https://doi.org/10.1037/cdp0000120

Wright C. V., Beattie S. G., Galper D. I., Church A. S., Bufka L. F., Brabender V. M., & Smith B. L. (2017). Assessment practices of professional psychologists: Results of a national survey. *Professional Psychology: Research and Practice*, *48*(2), 73–78. https://doi.org/10.1037/pro0000086