

Supplementary Table 1. Review of existing measures of cultural knowledge

◆ indicates recommended measure

Knowledge measure and usage	Sample items or techniques	Reliability/ validity	Pros	Cons
<i>Language knowledge</i>				
<p>Language proficiency scale (Luna & Peracchio, 2001).</p> <p>Used by Luna, Ringberg, & Peracchio (2008). Ringberg, Luna, Reihlen, & Peracchio (2010).</p>	<ul style="list-style-type: none"> ● “Rate how fluent you currently are in each of your languages (Spanish and English) in both Speaking and Listening (i.e., 4 items).” ● “Rate how well you think you can do the following things in English and Spanish (9 items): <ul style="list-style-type: none"> ○ understand cooking directions... ○ understand newspaper headlines ...” etc. ● 5-point Likert: 1 = very low/bad to 5 = like a native speaker/very well. 	<ul style="list-style-type: none"> ● Luna et al. (2001). Reliability: $\alpha = .94$ for Spanish items; .94 for English items. ● Luna et al. (2008). Predictive validity: Language items demonstrated higher scores for first versus second language. ● Ringberg, Luna, Reihlen, & Peracchio (2010). Reliability: $\alpha = .94$ Spanish items; .94 English items. 	<ul style="list-style-type: none"> ● Simple to administer. ● Assesses speaking, listening, reading and writing ability in separate items. ● Measures language fluency in different tasks (e.g., writing a letter, writing an advertisement). ● $\alpha = .94$. ● Evidence of predictive validity. ● Able to adapt questions to suit a particular language or tasks. 	<ul style="list-style-type: none"> ● Self-perceived language fluency may differ from objective language fluency, and may depend on domain. ● Tasks may be irrelevant in work context (e.g., ‘understand cooking directions’) or outdated (e.g., “writing a letter”).
<p>Survey questions assessing ‘Language usage’ as part of the Survey on the Vitality of Official-Language Minorities (SVOLM) collected by Statistics Canada.</p> <p>Used by Freynet & Clément (2015).</p>	<ul style="list-style-type: none"> ● 2 items assess perceived reading and writing skills for each of French and English, on a 5-point Likert scale ranging from “does not know how to read/write English/French” to “very good”. ● Participants considered bilingual if they scored high on both English and French language usage scales, based on a median split. 	<ul style="list-style-type: none"> ● Freynet & Clément (2015): Reliability: $\alpha = .88$ for Language usage 	<ul style="list-style-type: none"> ● Simple to administer. ● $\alpha = .73$ to .88 for language usage. 	<ul style="list-style-type: none"> ● Language usage may differ from objective language fluency, and may depend on domain. ● Only assesses reading and writing skills, not listening and speaking skills. ● Validity information not reported.

Knowledge measure and usage	Sample items or techniques	Reliability/ validity	Pros	Cons
<p>Language proficiency and usage (Benet-Martínez & Haritatos, 2005).</p> <p>Used by Chen et al. (2016; adapted).</p>	<ul style="list-style-type: none"> ● “Rate your overall Chinese/English language ability.” ● “How much do you use/have used Chinese/English to speak with your parents?” ● “How often do you watch TV shows/movies in Chinese/English?” 	<ul style="list-style-type: none"> ● Benet-Martínez & Haritatos (2005). Reliability: $\alpha = .82$ and $.85$ for English and Chinese language proficiency and usage scales, respectively. ● Chen et al. (2016). Reliability: $\alpha = .72$ and $.74$ for English and Chinese language proficiency and usage scales, respectively. 	<ul style="list-style-type: none"> ● Assesses language usage in a variety of domains. ● α for language proficiency = $.72$ to $.85$. 	<ul style="list-style-type: none"> ● Asking participants about overall language ability is not appropriate for individuals who have differing levels of speaking, listening, reading, and writing skills. ● Self-perceived language ability may differ from objective language fluency, and may depend on domain. ● Conflates language proficiency and language usage. ● Validity information not reported.
<p>Suinn-Lew Asian Self-Identity Acculturation Scale (SL-ASIA) - language items (Suinn, Rickard-Figueroa, Lew, & Vigil, 1987).</p> <p>Used by Dao, Teten, & Nguyen, (2011). Sirikantraporn, (2013).</p>	<ul style="list-style-type: none"> ● 4 language items: <ul style="list-style-type: none"> ○ “What language do you speak?” ○ “What language do you prefer?” ○ “Do you read only (Asian vs English language)?” ● “Do you write only (Asian vs English language)?” ● Score ranged from 1 (low acculturation) to 5 (high acculturation). 	<ul style="list-style-type: none"> ● Suinn et al. (1987). Predictive validity of overall scale inferred from significant ANOVA as a function of time lived in location. ● Dao et al. (2011). Convergent validity of overall scale: Cohen’s Kappa ranged from $.07$ to $.32$. Reliability of overall scale: $\alpha = .88$ for Asian sample; $.89$ for Chinese; $.88$ for Japanese; $.91$ for Korean; $.79$ for Filipino; and $.83$ for Vietnamese. ● Sirikantraporn (2013). Reliability and validity: information not reported. 	<ul style="list-style-type: none"> ● Scale contains other items that assess identity and internalization. ● Assesses speaking, reading, and writing in separate items. ● α for whole scale of 21 items = $.79$ to $.91$. ● Other validity information inferred. 	<ul style="list-style-type: none"> ● Assesses language usage and preference, not language proficiency. ● Does not specifically assess listening. ● Reliability and validity information is for the full SL-ASIA scale, not for the language items separately.

Knowledge measure and usage	Sample items or techniques	Reliability/ validity	Pros	Cons
<p>Acculturation Rating Scale for Mexican Americans-II (ARSMA-II) - language items (Cuéllar, Arnold, & Maldonado, 1995).</p> <p>Used by Basilio, Knight, O'Donnell, Roosa, Gonzales, Umaña-Taylor, & Torres (2014).</p>	<ul style="list-style-type: none"> ● Scale has 15 language items for English and Spanish, e.g., <ul style="list-style-type: none"> ○ “I speak Spanish/English”. ○ ”I enjoy Spanish/English language music”. ○ “I write letters in Spanish/English”. ○ “My thinking is done in Spanish/English”. 	<ul style="list-style-type: none"> ● Suinn et al. (1987). Reliability of overall scale: $\alpha = .88$. ● Cuéllar et al. (1995). Convergent validity: $r = .89$ between original ARSMA and ARSMA-II. Predictive validity: Anglo Orientation Score (AOS) pairwise comparison of five generation means were different, $F(4, 354) = 37.49, p < .00$. Mexican Orientation Score (MOS) yielded similar results, $F(4, 353) = 42.05, p < .001$. Reliability: $\alpha = .83$ for AOS; $.88$ for MOS; $.87$ for MOS Marginal. Test-retest reliability with a one-week interval was $.94$ for AOS; $.96$ for MOS; $.78$ for Marginal. ● Basilio et al. (2014). Reliability: English-language usage: $\alpha = .69$ for adolescents, $.91$ for mothers, $.89$ for fathers. Spanish-language usage: $\alpha = .83$ for adolescents, $.87$ for mothers, $.82$ for fathers. 	<ul style="list-style-type: none"> ● Full ARSMA-II scale also contains items that assess identity and internalization. ● Assesses self-perceived speaking, reading, writing, and listening proficiency in separate items. ● $\alpha = .83$ to $.88$ from orientation subscales. ● α for marginality subscales = $.68$ to $.91$. ● Evidence of predictive validity. 	<ul style="list-style-type: none"> ● 30 item scale to assess two languages leaves less time to assess other constructs. ● Conflates language proficiency with language preference and usage. ● Self-perceived language fluency could differ from objective language fluency. ● Does not specifically ask about language fluency in different domains (e.g., home, work).
<p>Language ability scale (Benet-Martínez, Lee, &</p>	<ul style="list-style-type: none"> ● 1 item per language: Chinese-American participants self-reported 	<ul style="list-style-type: none"> ● Not reported. 	<ul style="list-style-type: none"> ● Simple to administer. ● Face validity - it assesses language ability. 	<ul style="list-style-type: none"> ● Single-item scale does not capture speaking, listening, reading, and writing ability

Knowledge measure and usage	Sample items or techniques	Reliability/ validity	Pros	Cons
Leu, 2006).	fluency in Chinese and English languages on a scale from 1 to 5 (perfectly fluent).			separately. <ul style="list-style-type: none"> ● Self-perceived language fluency could differ from objective language fluency. ● Does not specifically ask about language fluency in different domains (e.g., home, work). ● Other reliability & validity information not reported.
Interview questions (Barker, 2015, 2017).	<ul style="list-style-type: none"> ● Barker, 2017. Participants were asked to describe their “fluency in the host country language and their cultural competency in general”. ● Barker, 2015. Participants were asked to describe their fluency in the host-country language. 	<ul style="list-style-type: none"> ● Barker (2015). Transparent sampling, data collection, and analysis processes; “member checking” of analysis 	<ul style="list-style-type: none"> ● Allows participants to describe, which can provide more nuanced insights than quantitative measures. ● Face validity - it assesses language fluency. 	<ul style="list-style-type: none"> ● Self-perceived language fluency could differ from objective language fluency. ● Does not specifically ask about language fluency in different domains (e.g., home, work).
Language and other knowledge				
<p>◆ Abbreviated Multidimensional Acculturation Scale (AMAS-ZABB) - language and cultural knowledge items (Zea, Asner-Self, Birman, & Buki, 2003).</p> <p>Used by Carrera & Wei (2014).</p>	<ul style="list-style-type: none"> ● 18 items measuring language proficiency, e.g., “How well do speak your native language: With family? On the phone? In general?” ● 12 items measuring cultural knowledge (called cultural competence), e.g., “How well do you know: National heroes from your native culture? Popular television shows in your native language? History of your native culture?” 	<ul style="list-style-type: none"> ● Zea et al. (2003). Concurrent validity: Significant difference between U.S. born vs. Latin American born. Hispanicism and Americanism were correlated with length of residence in the U.S., $r = -.24$ and $.21$, respectively. Convergent and discriminant validity of overall AMAS-ZABB scale: Positively correlated with a range of language and ethnic identity 	<ul style="list-style-type: none"> ● Scale measures both language and other types of cultural knowledge. ● Assesses self-perceived speaking and listening ability (understanding) in separate items. ● Assesses language ability in different tasks/domains. ● Able to adapt questions to suit a particular language or culture. ● AMAS-ZABB is well-validated for Latino samples. 	<ul style="list-style-type: none"> ● Only assesses self-perceived speaking and listening ability; does not assess reading or writing ability. ● 30-item scale (18 + 12) leaves less time to assess other constructs, although researchers could use only the language or cultural knowledge subscales. ● Validated for Latino sample; it may not be valid for other cultural groups.

Knowledge measure and usage	Sample items or techniques	Reliability/ validity	Pros	Cons
		<p>measures, such as Latino ethnic identity $r = .47$; Spanish language .46; Culture of origin .41; Multigroup Ethnic Identity Measure (MEIM) - Ethnic identity $r = .37$ Latino cultural identity, .32 English language. Negatively correlated with U.S.-American identity $r = -.45$, and overall U.S.-American dimension BIQ-B Hispanicism scale $-.36$.</p> <p>Construct validity: Proposed 6 subscales emerged as 6 factors, in multiple samples.</p> <p>Reliability: Values reported across study 1 and study 2 respectively: $\alpha = .96$ and $.97$ for English language; $\alpha = .90$ and $.89$ for U.S. cultural competence; $\alpha = .97$ and $.86$ for Spanish language; $\alpha = .93$ and $.83$ for Latino cultural competence.</p>		<ul style="list-style-type: none"> ● Cultural knowledge items focus on popular culture; scope may be too superficial and narrow as it does not assess deeper knowledge such as cultural values and beliefs.
<i>Knowledge other than language</i>				
<p>Bicultural Self-Efficacy Scale (BSES) (knowledge of cultural beliefs and values items) (David, Okazaki, & Saw, 2009).</p>	<ul style="list-style-type: none"> ● 4 items measuring knowledge of history, values, gender roles and celebrations, e.g., “I am knowledgeable about the values important to mainstream Americans as 	<ul style="list-style-type: none"> ● David et al. (2009). Construct validity: Six-factor solution fit better than one-factor and five-factor solutions. Concurrent validity: Study 1 - All six BSES subscales 	<ul style="list-style-type: none"> ● Assesses various types of knowledge. ● Relatively simple to administer. ● Able to adapt questions to suit a particular language or culture. 	<ul style="list-style-type: none"> ● Self-perceived cultural knowledge may differ from objective knowledge. ● Items are double-barrelled, limiting its utility for assessing hybrid cultures or three or more cultures.

Knowledge measure and usage	Sample items or techniques	Reliability/ validity	Pros	Cons
Used by Wei, Liao, Chao, Mallinckrodt, Tsai, & Botello-Zamarron (2010).	<p>well as to my cultural group”.</p> <ul style="list-style-type: none"> • Six subscales: Social Groundedness; Communication Ability; Positive Attitudes Towards Both Groups; Knowledge of Cultural Beliefs and Values; Role Repertoire; and Bicultural Beliefs. 	<p>correlated with the Mainstream and Heritage subscale of the Vancouver Index of Acculturation (VIA), $r = .15$ to $.65$. Five of the BSES subscales correlated with the Life Satisfaction Index, $r = .12$ to $.31$. Study 2 - BSES correlated with the Collective Self-Esteem Scale, $r = .18$ to $.50$.</p> <p>Convergent validity: Correlated with Collective Self-Esteem Scale, $r = .13$ to $.23$.</p> <p>Discriminant validity: Study 1 - BSES subscales were not strongly correlated with a number of the Mood and Anxiety Symptoms Questionnaire (MASQ) subscales, $r = -.10$ to $-.30$. Study 2 - BSES subscales significantly correlated with The Balanced Inventory of Desirable Responding - Self-Enhancement subscale, $r = .13$ to $.23$; but not correlated with the Impression Management or Self-Denial subscale.</p> <p>Reliability: Reliabilities for each of the six subscales $\alpha = .63$ to $.93$ across two studies.</p> <ul style="list-style-type: none"> • Wei et al. (2010). <p>Reliability: $\alpha = .92$ for overall</p>	<ul style="list-style-type: none"> • Some evidence of concurrent, convergent and discriminant validity. 	<ul style="list-style-type: none"> • Strong reliability scores for total scale, mixed results for subscales. • Mixed results for validity.

Knowledge measure and usage	Sample items or techniques	Reliability/ validity	Pros	Cons
		scale; .46 to .89 for the subscales.		
Ethnic and British knowledge measures (as part of bicultural competence items) (Benbow & Rutland, 2017)	<ul style="list-style-type: none"> Scale items include knowledge of food, history, music and values. 	<ul style="list-style-type: none"> Reliability: $\alpha = .88$ for ethnic knowledge, $.85$ for British knowledge. No validity information. 	<ul style="list-style-type: none"> Assesses various types of knowledge. Relatively simple to administer. Able to adapt questions to suit a particular language or culture. 	<ul style="list-style-type: none"> Self-perceived cultural knowledge may differ from objective knowledge. Ethnic knowledge relates to “the general group you and your parents belong to rather than your individual ethnic identity,” which is ambiguous. No validity information.
Narrative inquiry (Kanno, 2000).	<ul style="list-style-type: none"> 2-3 interviews before participants left Canada, then 6-monthly visits following their return to Japan over two years. 	<ul style="list-style-type: none"> This technique is appropriate for interpretive research designs. Transparency will vary across projects, depending how it is applied. 	<ul style="list-style-type: none"> Rich data where reliability is supported by reporting participants’ own words wherever possible. 	<ul style="list-style-type: none"> Time consuming. This technique is especially appropriate for understanding lived experiences. It may be more challenging to use it to assess multicultural knowledge.
Critical ethnography (Peñaloza & Gilly, 1999).	<ul style="list-style-type: none"> Interviews and observations. 	<ul style="list-style-type: none"> This technique is appropriate for interpretive research designs. Transparency will vary across projects, depending how it is applied. 	<ul style="list-style-type: none"> Rich data. Technique may be useful for examining participants’ situated knowledge within context, such as within organizational contexts. 	<ul style="list-style-type: none"> Time consuming. Technique is especially appropriate for examining implicit values and hidden biases, which is theoretically closer to internalization than knowledge.

Supplementary Table 2. Review of existing measures of cultural identification

◆ indicates recommended measure

Identification measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
<i>Categorical – Single item, culture-general</i>				
Single item with description (closed answer) (Fitzsimmons, Liao, & Thomas, 2017).	<ul style="list-style-type: none"> ◆ “Do you have more than one cultural identity? A cultural identity is a culture that is so familiar to you that it becomes part of who you are. A culture can refer to a region or a country. For example, Chinese, Indian, and French-Canadian are all cultures. You can be a member of a culture even if you’ve never lived there, but it must be so deeply embedded in you that it influences your values, your behaviors and the way you see the world”. 	<ul style="list-style-type: none"> ● Not reported. 	<ul style="list-style-type: none"> ● Simple method to differentiate monoculturals from multicultural. ● Allows individuals to self-identify (i.e. identities are not imposed upon respondents). ● Inclusion of a description of cultural identity reduces variation in interpretation of the term. 	<ul style="list-style-type: none"> ● Description conflates identification with internalization. ● Description does not capture multidimensionality of identity. ● Creates a dichotomy (monocultural and multicultural) rather than treating multiculturalism as a continuum. ● Participants may not read long description. ● Reliability and validity information not reported.
Single item with description (open answer) (Hoersting & Jenkins, 2011).	<ul style="list-style-type: none"> ◆ “Most people belong to one or more social groups, communities, or networks. There might be several ways to describe people who have had a variety of cross-cultural experiences. Is there a particular label that you feel best describes a group that encompasses your childhood cross-cultural experience? If so, what is that?” Prompted with labels such as third culture kids, global nomad, or military brat (open-ended 	<ul style="list-style-type: none"> ● Not reported. 	<ul style="list-style-type: none"> ● Allows individuals to self-identify (i.e. identities are not imposed upon respondents). 	<ul style="list-style-type: none"> ● Conflates experience with identification. ● Creates a dichotomy (monocultural and multicultural) rather than treating multiculturalism as a continuum. ● Participants may not read long description. ● Coding of open-ended responses may be difficult and time-consuming. ● Reliability & validity information not reported.

Identification measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
	response, coded dichotomously).			
Single item without description. (Benet-Martínez & Haritatos, 2005).	<ul style="list-style-type: none"> • “How much do you identify with U.S. (Chinese) culture?” . 6-point Likert ranges from 1 = very weakly identified to 6 = highly identified. 	<ul style="list-style-type: none"> • Not reported. 	<ul style="list-style-type: none"> • Simple method to differentiate monoculturals from multicultural. • Allows individuals to self-identify (i.e. identities are not imposed upon respondents). 	<ul style="list-style-type: none"> • Creates a dichotomy (monocultural and multicultural) rather than treating multiculturalism as a continuum. • Does not capture multidimensionality of identity. • Reliability & validity information not reported.
<i>Categorical – Single item, culture-specific</i>				
Various: Mok & Morris (2013). Perunovic, Heller, & Rafaeli (2007). Ward (2006). Ting-Toomey, Yee-Jung, Shapiro, Garcia, Wright, & Oetzel (2000). Rivera-Sinclair (1997).	<ul style="list-style-type: none"> • “I identify with both American and East Asian culture” (Mok & Morris, 2013). • “During the past 2 hours, which specific cultural group did you most identify with?” Perunovic et al. (2007). • “What is your ethnic background?: New Zealand European, Maori, Other (please specify)” (Ward, 2006). • Participants self-identified as European American, African American, Asian American, or Latino(a) American (Ting-Toomey, et al., 2000). • Self-identified; Cuban vs. Cuban-American (Rivera-Sinclair, 1997). 	<ul style="list-style-type: none"> • Not reported. 	<ul style="list-style-type: none"> • Most of these examples are simple methods to differentiate monoculturals from multicultural. • Face validity - participants self-identified. 	<ul style="list-style-type: none"> • Creates a dichotomy (monocultural and multicultural) rather than treating multiculturalism as a continuum. • Pre-defined categories do not allow for identification with a hybrid culture. • Some approaches restrict the potential to identify highly with more than one culture (e.g. Perunovic et al., 2007) • Demographic and ethnicity measures do not capture identification • Does not capture multidimensionality of identity. • Reliability & validity information not reported.
Immigrant	<ul style="list-style-type: none"> • First generation immigrant 	<ul style="list-style-type: none"> • Not reported 	<ul style="list-style-type: none"> • Immigrant generation variables 	<ul style="list-style-type: none"> • Immigrant generation is

Identification measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
<p>Generation</p> <p>Used by Hsu (2011). Benet-Martínez, Lee & Leu (2006).</p>	<p>(Benet-Martínez et al. (2006).</p> <ul style="list-style-type: none"> • Second generation immigrant (Hsu, 2011). 		<p>may be more available than identification variables for projects using secondary data.</p>	<p>unlikely to be a reliable proxy for multicultural identification.</p>
<p>Suinn-Lew Asian Self-Identity Acculturation Scale (Suinn, Rickard-Figueroa, Lew, & Vigil, 1987).</p> <p>Used by Chiao et al. (2010). Dao, Teten, & Nguyen (2011). Sirikantraporn, (2013).</p>	<ul style="list-style-type: none"> • Items in this scale related to identity, e.g.: “How would you rate yourself? 1. Very Asian 2. Mostly Asian 3. Bicultural 4. Mostly Westernized 5. Very Westernized” (Suinn et al., 1987). 	<ul style="list-style-type: none"> • Not reported. 	<ul style="list-style-type: none"> • Simple method to differentiate monoculturals from multicultural. • Face validity - participants self-identified. 	<ul style="list-style-type: none"> • Creates groups rather than treating multiculturalism as a continuum. • Cultures may be considered too broad (e.g., Asian rather than Chinese). • Question format means this scale can only be used to assess identification with two cultures, not more • Pre-defined categories do not allow for identification with a hybrid culture. • Does not capture multidimensionality of identity. • Reliability & validity information not reported.
<p>Ethnic self-categorization (Kulis, Robbins, Baker, Denetsois, & Deschine Parkhurst, 2016).</p>	<ul style="list-style-type: none"> • Participants were asked if they considered themselves “American Indian only”, “an equal member of multiple ethnic or racial groups”, “more American Indian than other ethnicities”, “more a member of (another ethnic/racial) group than American Indian”, or not “a part of any of these groups”. This item was dichotomized (American 	<ul style="list-style-type: none"> • Not reported. 	<ul style="list-style-type: none"> • Single-item scale is easy to administer. • Response set allows for the possibility that participants identify primarily with another culture, beyond American Indian culture. 	<ul style="list-style-type: none"> • Creates distinct groups rather than treating multiculturalism as a continuum. • Question format means this scale can only be used to assess identification with two cultures, not more.

Identification measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
Indian versus other responses) in the latent class analysis.				
<i>Continuous – Single item</i>				
<p>Various: Huff, Lee, & Hong (2017). Hanek, Lee, & Brannen (2014). Carpentier & de la Sablonniere, (2013). Saad, Damian, Benet-Martínez, Moons, Robins (2013). Mok & Morris (2013; Study 1). Mok & Morris (2010ab). Chen, Benet-Martínez, & Bond (2008). Cheng, Lee, & Benet-Martínez (2006). Benet-Martínez & Karakitapoğlu-Aygün (2003). Benet-Martínez, Lee, Leu, & Morris (2002).</p>	<ul style="list-style-type: none"> ● Level of identification with Chinese [and American] culture (Benet-Martínez et al., 2006). ● “I feel the Chinese identity in me is...” (Cheng et al., 2006). ● “To what extent do you identify with this country?” (Hanek et al., 2014). ● “How much do you identify with Hong Kong culture?” (Chen et al., 2008). ● “I identify with members of my original group” (Carpentier & de la Sablonniere, 2013). 	<ul style="list-style-type: none"> ● Not reported. 	<ul style="list-style-type: none"> ● Simple method. ● It allows individuals to self-identify (i.e. identities are not imposed upon respondents). ● Identification is on a continuum (i.e. not dichotomous). 	<ul style="list-style-type: none"> ● Cultures may be considered too broad (e.g. “East Asian” per Mok & Morris, 2010a) ● Typically, pre-defined categories do not allow for identification with a hybrid culture. ● Single item does not capture multidimensionality of identity.
<i>Continuous – Single item visual</i>				
<p>Inclusion of Ingroup in the Self (IIS) (Tropp & Wright, 2001).</p>	<ul style="list-style-type: none"> ● A single visual item displaying the relationship between two circles – one representing the individual, 	<ul style="list-style-type: none"> ● Tropp & Wright (2001). Reliability: Test-retest reliability = .76. Convergent validity: Correlated 	<ul style="list-style-type: none"> ● Simple, creative method allows individual expression of identity. ● Allows individuals to free-list 	<ul style="list-style-type: none"> ● Difficult to compare across individuals (e.g., if someone lists Chinese-American versus if someone lists

Identification measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
Used by Yampolsky, Amiot, & de la Sablonniere (2016).	and the other representing their cultural group – and rated on a scale ranging from 1 (completely separate circles) to 7 (completely overlapping circles).	with “integration” subscale of MULTIIS (Yampolsky et al., 2016).	<p>their identities (i.e. identities are not imposed upon respondents).</p> <ul style="list-style-type: none"> • Identification is on a continuum (i.e. not dichotomous). • Evidence of test-retest reliability and convergent validity with MULTIIS. 	<p>Chinese and American).</p> <ul style="list-style-type: none"> • Does not capture multidimensionality of identity.
<i>Continuous – Multi-item scale</i>				
<p>Multigroup Ethnic Identity Measure (MEIM) (Phinney, 1992); revised version (MEIM-R) presented by Phinney and Ong (2007)</p> <p>Used by Benbow & Rutland (2017). Spiegler & Leyendecker (2017). Kulis, Robbins, Baker, Denetsosie, & Deschine (2016; adapted). Schwartz et al., (2015). Hoersting & Jenkins (2011; adapted). David, Okazaki, & Saw (2009). Gong (2007). Ward (2006; adapted). Phinney & Devich-</p>	<ul style="list-style-type: none"> • MEIM-R includes 6 items that assess two dimensions: <ul style="list-style-type: none"> ○ Identity exploration (3 items), e.g. “I have spent time trying to find out more about my ethnic group, such as its history, traditions and customs”. ○ Identity commitment (3 items), e.g. “I have a strong sense of belonging to my own ethnic group”. 	<ul style="list-style-type: none"> • Benbow & Rutland (2016). Reliability: $\alpha = .90$ • Kulis et al. (2016). Reliability: $\alpha = .90$. • Phinney & Devich-Navarro, (1997). Reliability: $\alpha = .75$. • Schwartz et al. (2015). Reliability: $\alpha = .85$ to $.92$. • Hoersting & Jenkins (2011). Reliability: $\alpha = .90$ for 7-item “affirmation” subscale of MEIM. • David et al. (2009). Reliability: $\alpha = .91$. • Gong (2007). Reliability: $\alpha = .87$ Construct Validity: 3-factor structure as hypothesized. • Ward (2006). Reliability: $\alpha = .70$ to $.86$. 	<ul style="list-style-type: none"> • Assesses two dimensions of identity. • Identification is on a continuum (i.e., not dichotomous). • Culture-general: can be used with all ethnic groups. • $\alpha = .70$ to $.92$. • Evidence of reliability and validity, especially for the revised measure MEIM-R (Phinney & Ong, 2007). 	<ul style="list-style-type: none"> • Emphasis is on ethnic or heritage identity, not cultural identity. May be less appropriate for assessing identification with cultures not linked to ethnicity.

Identification measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
Navarro (1997).				
<p>Collective Self-Esteem Scale (CSES) (Luhtanen & Crocker, 1992).</p> <p>Used by Wiley (2013; adapted). David, Okazaki, & Saw (2009). Downie, Mageau, Koestner, & Liodden (2006). Devos (2006).</p>	<ul style="list-style-type: none"> Membership subscale, e.g., “I am a worthy member of the social groups I belong to”. Private subscale, e.g., “In general, I’m glad to be a member of the social groups I belong to”. Public subscale, e.g., “In general, others respect the social groups I am a member of.” Identity subscale, e.g., “The social groups I belong to are an important reflection of who I am”. 	<ul style="list-style-type: none"> Luhtanen & Crocker (1992). Reliability: $\alpha = .83$ to .88. Construct validity: Four-factor structure as hypothesized. David et al. (2009). Reliability: $\alpha = .48$ to .87 for four subscales. Convergent validity: correlated with BSES. Devos (2006). Reliability: $\alpha = .79$ to .95 for identity subscale. Downie et al. (2006). Reliability: $\alpha = .84$ 	<ul style="list-style-type: none"> Allows individuals to self-identify (i.e. identities are not imposed upon respondents). Identification is on a continuum (i.e. not dichotomous). Theoretically derived identity dimensions. Evidence of reliability and validity. 	<ul style="list-style-type: none"> 16-item scale leaves less time to assess other constructs. Self-esteem is not cultural identification.
<p>Abbreviated Multidimensional Acculturation Scale (AMAS-ZABB) - identity items (Zea, Asner-Self, Birman, & Buki, 2003).</p> <p>Used by Sandil, Robinson, Brewster, Wong, & Geiger (2015). Carrera & Wei (2014).</p>	<ul style="list-style-type: none"> 6 items measuring American identity, e.g., “I think of myself as being American”; 6 items measuring ethnic identity, e.g., “I have a strong sense of being [<i>ethnic group</i>]”. 	<ul style="list-style-type: none"> Zea et al. (2003). Reliability: $\alpha = .89$ to .96 (Identity items). Convergent and discriminant validity: AMAS-ZABB U.S.- American identity showed positive and negative correlations with other measures as expected, such as with Bicultural Inventory Questionnaire - Form B (BIQ-B) Americanism $r = .40$; English language .48; U.S.- American cultural competence .31; Latino identity .01; Spanish language -.16; Latino cultural competence -.22. Construct validity: Proposed 6 subscales emerged as 6 factors 	<ul style="list-style-type: none"> Allows individuals to self-identify (i.e. identities are not imposed upon respondents). Identification is on a continuum (i.e. not dichotomous). Full AMAS-ZABB scale also includes knowledge items. Evidence of reliability and validity. Scale items could be adapted for use with cultural identities beyond ethnicity. 	<ul style="list-style-type: none"> Does not draw on theory to assess dimensions of identity. Designed to assess identification with an ethnic group and “mainstream” cultural group,

Identification measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
		<p>in multiple samples.</p> <ul style="list-style-type: none"> • Sandil et al. (2015). Reliability: $\alpha = .89$ to $.94$. • Carrera & Wei (2014). Reliability: $\alpha = .91$ to $.93$. 		
<p>Revised Acculturation Rating Scale for Mexican Americans (ARSMA-II) (Cuéllar, Arnold, & Maldonado, 1995).</p> <p>Used by Ferguson, Ferguson, & Ferguson (2017). Cuéllar, Arnold, & Maldonado (1995).</p>	<ul style="list-style-type: none"> • “I like to identify myself as a Mexican American”. 	<ul style="list-style-type: none"> • Cuéllar et al. (1995). Reliability: $\alpha = .83$ to $.88$ Concurrent Validity: ARSMA-II correlated with ARSMA $r = .89$. Convergent Validity: ARSMA-II correlated with generational status $r = .61$. ARSMA-II means significantly different across five different levels of generational status ($F=54.2$, $p<0.001$). All but one of the pairwise comparisons are significantly different. • Ferguson et al. (2017). Reliability: $\alpha = .78$ to $.89$. 	<ul style="list-style-type: none"> • Possibly could be adapted to different cultures. • ARSMA-II is based on an orthogonal approach to acculturation, which is better than a linear approach. 	<ul style="list-style-type: none"> • Scale is meant for two cultures only. • Scale was developed for Mexican Americans; it may not be valid for other groups. • “Anglo,” may tap multiple cultures; it may fit ethnicity better than culture. • Validity information not reported.
<p>◆ Social identity scale (Roccas, Sagiv, Schwartz, Halevy, & Eidelson, 2008).</p> <p>Used by Fitzsimmons et al. (2017; used identity importance subscale only).</p>	<ul style="list-style-type: none"> • Importance subscale, e.g. “It is important to me that I view myself as a member of this group”. • Commitment subscale, e.g., “I feel strongly affiliated with this group”. • Superiority subscale, e.g., “This group is better than other groups in all respects”. • Deference subscale, e.g., “It is disloyal to criticize this group”. 	<ul style="list-style-type: none"> • Roccas et al. (2008). Construct validity: CFAs with samples from the U.S. and Israel support model of four distinct but related factors. Multigroup CFA supported model fit across samples and over time, with invariant factor loadings and covariances; subscales were intercorrelated by $.39$ to $.79$. Predictive validity: Patterns of correlations with personality traits. • Fitzsimmons et al. (2017). 	<ul style="list-style-type: none"> • Allows individuals to self-identify (i.e. identities are not imposed upon respondents). • Identification is on a continuum (i.e. not dichotomous). • Theoretically derived identity dimensions. • $\alpha = .80$ to $.94$. • Evidence of validity through multigroup CFA model fit across international samples and over time, with invariant factor loadings and covariances. • Validity was further established 	<ul style="list-style-type: none"> • 16-item scale leaves less time to assess other constructs, though it can be divided into theoretically distinct sub-scales, such as the 4-item identity importance sub-scale. • No evidence of criterion validity.

Identification measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
		<p>Validity: The identity importance subscale performed as expected against a series of one-way ANOVAs with groups of monoculturals, biculturals and multicultural.</p> <p>Reliability: $\alpha = .80$ to $.94$ for importance subscale.</p>	<p>by examining relationships with stable constructs (personality) and dynamic constructs (context effects).</p> <ul style="list-style-type: none"> ● Identity importance subscale performed as expected against a series of one-way ANOVAs with groups of monoculturals, biculturals and multicultural. ● Scale can be adapted for use with most cultural groups. 	
<p>◆ In-group identification Cameron (2004).</p> <p>Used by Stroink & Lalonde (2009).</p>	<ul style="list-style-type: none"> ● 12 item scale across 3 dimensions: <ul style="list-style-type: none"> ○ Ingroup Ties, e.g., "I feel strong ties to other (ingroup members)". ○ Centrality, e.g., "In general, being a(n) (ingroup member) is an important part of my self-image". ○ Ingroup Affect, e.g. "In general, I'm glad to be a(n) (ingroup member)". 	<ul style="list-style-type: none"> ● Cameron (2004). Construct validity: CFA results across five samples totalling 1078 participants support the three-factor model over alternatives. All three factors were significantly related to each other, ranging from $.20$ to $.61$. ● Reliability: Test-retest reliability one week apart was supported with correlations from $.65$ to $.77$. ● Stroink & Lalonde (2009). Reliability: $\alpha = .77$ to $.85$. 	<ul style="list-style-type: none"> ● Allows individuals to self-identify (i.e. identities are not imposed upon respondents). ● Identification is on a continuum (i.e. not dichotomous). ● Theoretically derived and empirically supported identity dimensions. ● Captures affective and interpersonal components of identity that are sometimes ignored by other measures. ● Evidence of reliability and validity across various samples. 	<ul style="list-style-type: none"> ● Strong relationship and significant chi-square scores between ingroup ties and ingroup affect indicate these may not always be distinct.
<p>Local-global identity scale (Zhang & Khare, 2009).</p>	<ul style="list-style-type: none"> ● "I believe I mostly belong to my local community". ● "I believe that people all over the world are more similar than different". 	<ul style="list-style-type: none"> ● Reliability: $\alpha = .63$ to $.70$. Individuals who identified more with a global identity preferred global products, while the opposite was true for individuals who identified more with a local identity. 	<ul style="list-style-type: none"> ● Identification is on a continuum (i.e., not dichotomous). ● Scale has some early evidence of predictive validity, as individuals who identified more with a global identity preferred global products, while the opposite was true for individuals who identified more with a local identity (Zhang & 	<ul style="list-style-type: none"> ● 19-item scale leaves less time to assess other constructs. ● Conflates identity with beliefs and attitudes. ● Specific to global and local identities. ● Items are not all parallel with respect to global and local identities, with different

Identification measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
			Khare, 2009).	<p>questions for each scale, and one additional item assessing local identity.</p> <ul style="list-style-type: none"> ● Mixed evidence for scale reliability, with α ranging from .63 to .70.
<p>Jewish-American Identity Scale (Zak, 1973).</p> <p>Used by Der-Karabetian & Ruiz (1997).</p>	<ul style="list-style-type: none"> ● American identity, e.g., “Being an American plays an important part in my life”. ● Jewish identity, e.g., “My fate and future are bound up with Jews everywhere”. 	<ul style="list-style-type: none"> ● Der-Karabetian & Ruiz (1997). <p>Reliability: $\alpha = .70$ to .81 for adapted version of scale.</p> <p>Construct validity: No correlation between Latino/a and American identities; both were correlated with “global-human” identity.</p> <p>Predictive validity: The children of American immigrants scored higher on American identity than first-generation immigrants; there was no difference on Latino/a identity across immigrant generations.</p>	<ul style="list-style-type: none"> ● Scale designed to be tailored to one demographic group and was developed in consultation with group members. 	<ul style="list-style-type: none"> ● Original 19-item scale leaves less time to assess other constructs, while shortened scale by Der-Karabetian & Ruiz, 1997 is too different to be considered the same scale. ● Specific to Jewish-American identity; may not be valid for other cultures. ● Not theory-based.
<p>Ethnic/Cultural Identity Dimensions Scale (Ting-Toomey, Yee-Jung, Shapiro, Garcia, Wright & Oetzel, 2000).</p>	<p>49 item scale across 4 dimensions:</p> <ul style="list-style-type: none"> ● Belonging (15 items), e.g., “I have spent time trying to find out more about my own ethnic group, such as history, traditions and customs”. ● Fringe (11 items), e.g., “I feel like I live on the ‘fringe’ in terms of my sense of ethnic group belongingness.” ● Interaction (14 items), e.g., 	<ul style="list-style-type: none"> ● Ting-Toomey et al (2000). <p>Construct validity: Scale was developed by administering 84 items from related scales, applying an exploratory factor analysis with varimax rotation, and selecting items with a primary factor loading of at least .50. No confirmatory validation tests were completed.</p> <p>Reliability: $\alpha = .76$ to .92.</p>	<ul style="list-style-type: none"> ● Identification is on a continuum (i.e., not dichotomous). ● Multiple identity dimensions included in the scale. ● Scale developers attempted to fit factors into constructs suggested by acculturation research. ● $\alpha = .76$ to .92. 	<ul style="list-style-type: none"> ● 49-item scale leaves less time to assess other constructs. ● Conflates identity with behavior. ● Scale developed without strong theoretical justification or evidence of scale validity beyond exploratory factor analysis. ● Designed to assess identification with an ethnic group and “mainstream” cultural group; could be

Identification measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
	<p>“I feel unable to involve myself in activities with members of the other ethnic group(s)”.</p> <ul style="list-style-type: none"> Assimilation (9 items), e.g., “The values of my own ethnic group are very compatible with that of the overall US culture”. 			<p>difficult to adapt to other cultures.</p>
<p>Asian American Multidimensional Acculturation Scale (AAMAS) - identity subscale (Chung, Kim, & Abreu, 2004).</p> <p>Used by Lee & Church (2017). Gong (2007).</p>	<ul style="list-style-type: none"> “How much do you feel you have in common with Asian Americans?” “How much do you identify with Asian Americans?” 	<ul style="list-style-type: none"> Chung, Kim, & Abreu (2004). Reliability: Three studies supported scale reliability. Test-retest coefficients = .75 to .89 after a two-week lag. $\alpha = .76$ to .91. Predictive validity: Increased generational status predicted lower identification with culture of origin; it had no relationship Asian American or European American identification. Convergent validity: Scale was related to other cultural identity scales. Divergent validity: supported by no relationship with self-esteem, although it had a small positive relationship to intergenerational conflict. Gong, 2007. Reliability: $\alpha = .76$ to .82. Church, 2017. Reliability: $\alpha = .76$ to .91. 	<ul style="list-style-type: none"> Evidence of both reliability and validity across scale development studies and in further use by other researchers (e.g. Gong, 2007). $\alpha = .76$ to .91. Designed to be applicable across most cultures. This scale is distinct from others by measuring identification with a hybrid culture (Asian American). 6-item identity subscale is concise. Full scale also contains items to assess language and cultural knowledge. Identification is on a continuum (i.e., not dichotomous). 	<ul style="list-style-type: none"> Does not have subscales for different identity dimensions. Items combine behavioral acculturation and identity into the same subscale. Developed specifically for Asian American samples; may not be generalizable to other multicultural samples.

Identification measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
Cultural Identity List (Novin, Banerjee, & Rieffe, 2012).	<ul style="list-style-type: none"> ● “I feel at home in the Dutch/Moroccan culture”. ● “I have many Dutch/Moroccan friends”. 	<ul style="list-style-type: none"> ● Novin et al (2012). Reliability: $\alpha = .79$ to $.82$. ● Validity: Scale was developed by listing an unspecified number of items from other similar scales, with no attempt to validate. 	<ul style="list-style-type: none"> ● Identification is on a continuum (i.e., not dichotomous). ● $\alpha = .79$ to $.82$. 	<ul style="list-style-type: none"> ● No evidence of scale validity. ● Number of items is not stated. ● Pre-defined categories do not allow for identification with a hybrid culture. ● Conflates friendships and comfort with a culture with identification.
Ethnic Identity Scale (Umaña-Taylor, Yazedjian, & Bámaca-Gómez, 2004). Used by Basilio et al. (2014).	<ul style="list-style-type: none"> ● Exploration subscale (23 items), e.g., “I have attended events that have helped me learn more about my ethnicity”. ● Resolution subscale (13 items), e.g., “I understand how I feel about my ethnicity”. ● Affirmation subscale (10 items), e.g., “I have positive feelings about my ethnicity”. 	<ul style="list-style-type: none"> ● Umaña-Taylor, Yazedjian, & Bámaca-Gómez (2004). Reliability: $\alpha = .86$ to $.92$. ● Exploration and resolution subscales (but not affirmation) were positively related to self-esteem and familial ethnic socialization among ethnic minority group members. ● Basilio et al. (2014). Reliability: $\alpha = .81$ to $.84$. ● Ethnic identity was positively related to Mexican-American biculturalism. 	<ul style="list-style-type: none"> ● Theoretically derived identity dimensions. ● Could be applied to other cultures due to culture-general wording. ● $\alpha = .86$ to $.92$. 	<ul style="list-style-type: none"> ● Long scale (23 + 13 + 10) leaves less time to assess other constructs. ● Emphasis is on ethnicity and the process of identity formation. May be less appropriate for assessing identification with cultures not linked to ethnicity. ● Exploratory and confirmatory factor analyses were conducted on the same sample. More evidence is needed to support this scale’s construct validity.
Bulgarian Mainstream Identity Scale (Dimitrova, Chaslotis, Bender & van de Vijver, 2013). Used by van de Vijver, Blommaert, Gkoumasi, & Stogianni (2015);	<ul style="list-style-type: none"> ● “I consider myself Bulgarian”. ● “I participate in Bulgarian practices”. 	<ul style="list-style-type: none"> ● Dimitrova et al. (2013). Reliability: $\alpha = .90$ to $.93$. ● Validity information not reported. ● van de Vijver et al. (2015). Reliability: $\alpha = .89$. Authors removed more than half the original items and did not report validity information for new scale. 	<ul style="list-style-type: none"> ● Scale developers created items in an attempt to capture self-categorization, attachment, evaluation, importance and behavioral involvement. ● Identification is on a continuum (i.e., not dichotomous). ● $\alpha = .89$ to $.93$. 	<ul style="list-style-type: none"> ● Neither the original 21-item scale (Dimitrova et al., 2013), nor the 10-item adapted version (van de Vijver et al., 2015) were validated, except reporting acceptable α. ● Conflates identity with behavior. ● Original scale was specifically developed for

Identification measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
adapted).				the Bulgarian context; it may not be valid for other cultures.
Cultural identity (Hanek, Lee, & Brannen, 2014).	3-item scale: <ul style="list-style-type: none"> ● “To what extent do you identify with this country?” ● “To what extent are your values the same as this country?” ● “To what extent is your worldview the same as this country?” 	Reliability: $\alpha = .73$ for identification with home culture, $\alpha = .80$ for identification with host culture. No validity information reported.	<ul style="list-style-type: none"> ● Three items were selected based on prior evidence that they correlate with longer measures of cultural identification. ● 3-item scale is simple to administer. ● Could be applied to other cultures due to culture-general wording. 	<ul style="list-style-type: none"> ● Purportedly a measure of cultural identity, but two out of three scale items are about internalization – it conflates identity and internalization. ● Use of the term “country” implies homogenous national-level culture. ● No validity information was reported.
Measures of Markstrom’s (2011) model of local identity and national identity (Kulis, Robbins, Baker, Denetsose, Deschine & Parkhurst, 2016).	<ul style="list-style-type: none"> ● A battery of measures of the three components of Markstrom’s (2011) model of local identity: <ul style="list-style-type: none"> ○ Strength of Identification as American Indian. ○ Connections with American Indians. ○ Cultural/spiritual practices, language world view and values. ● Markstrom’s (2011) national level of identity was assessed with measures of bicultural orientations. 	<ul style="list-style-type: none"> ● Kulis et al. (2016) Reliability: not reported as each of the three components is operationalized with more than one measure, which are not expected to correlate. For example, cultural and spiritual practices were assessed by frequency of using a cultural language, and a six-item scale assessing whether participants follow the “American Indian” or “white / Anglo” way of life. Validity: supported through latent class analysis (LCA), where authors identified five clusters of individuals with similar patterns of responses. Although there was no statistical difference	<ul style="list-style-type: none"> ● Items were developed based on an emic theoretical model specific to Indigenous Americans. ● Five latent classes of individuals based on patterns of results were validated against open-ended identity descriptions. ● Extensive measures go beyond Likert-type agreement with statements. ● Designed to capture identification with a specific hybrid culture that is assumed to exist at least partially through blood ancestry; it should not be applied to other groups. 	<ul style="list-style-type: none"> ● The approach of using a battery of uncorrelated measures to assess each component rather than selecting only one measure is reasonable because they are theoretically derived. However, this approach makes it difficult to assess reliability. ● Some measures conflate identity and internalization/behavior.

Identification measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
		between the four- or five-cluster solutions, the five-cluster solution was selected because responses seem to correspond to their open-ended identity descriptions.		
Ethnic and English identity (Jugert et al., 2017)	<ul style="list-style-type: none"> 4-item measure, e.g., “How proud are you about being [ethnic group]?” 	<ul style="list-style-type: none"> Jugert et al. (2017) Reliability: $\alpha = .62$ to $.79$ for ethnic majority and $.71$ to $.73$ for ethnic minority children on the ethnic identity scale. English identification ranged from $.69$ to $.76$ for ethnic majority and $.82$ to $.84$ for ethnic minorities. 	<ul style="list-style-type: none"> 4-item scale is simple to administer. Identification is on a continuum (not dichotomous). Culture-general in terms of country and ethnicity. Some evidence of reliability. 	<ul style="list-style-type: none"> Designed to assess identification with an ethnic group and “mainstream” cultural group; could be difficult to adapt to other cultures Pre-defined categories do not allow for identification with a hybrid culture No validity information provided.
Qualitative methods				
Interview examples: Dey, Balmer, Pandit, Saren, & Binsardi (2017). Kane & Levina (2017). Barker (2015). Cross & Gilly (2014). Tawa & Suyemoto (2010). Friedlander, Larney, Skau, Hotaling, Cutting, & Schwam (2000).	<ul style="list-style-type: none"> “How do you understand yourself?” (Tawa & Suyemoto, 2010). Participants were asked to describe their cultural identities (Barker, 2015). “Do you feel proud to be American?” “Do you feel proud to be of Russian origin?” (Kane & Levina, 2017). 	<ul style="list-style-type: none"> Measures are supported by interpreting and prioritizing participants’ own voices and experiences. 	<ul style="list-style-type: none"> Interviews allow individuals to self-identify (i.e. identities are not imposed upon respondents) Data collected by interviews allow for novel discoveries not anticipated by researchers in advance. 	<ul style="list-style-type: none"> Open-ended responses can be difficult to interpret. Some individuals find it difficult to describe their cultural identities.
Twenty Statements Test (Kuhn & McPartland, 1954).	<ul style="list-style-type: none"> Participants listed 20 self-descriptions in response to the question, “Who am I?” 	<ul style="list-style-type: none"> Kuhn & McPartland (1954). Validity: “pragmatic test of the usefulness of the scale scores 	<ul style="list-style-type: none"> Allows participants to describe themselves in their own words. Well-established measure for a 	<ul style="list-style-type: none"> Coding is time-consuming. Validity was not reported in Ferguson et al., 2017, which is

Used by Ferguson, Ferguson, & Ferguson (2017).	<p>of the ‘locus’ component of self-attitudes may serve also as the second kind of demonstration of the validity of the instrument” (p. 73). Reliability: Coefficient of reproducibility = .90. • Ferguson et al. (2017). Validity information not reported. Inter-rater reliability ranged from .86 to .98 across each of the 10 response sets.</p>	<p>range of self-perceptions questions (i.e., not specific to cultural identity). • Evidence of reliability and validity.</p>	the measure most relevant to culture.	
Narrative inquiry examples: Kanno (2000). Oswald (1999). Thompson & Tambyah (1999). Yampolsky, Amiot & de la Sablonniere (2013).	<ul style="list-style-type: none"> • Participants engaged in a process of “narrative inquiry”, being interviewed many times during a two-year process of identity change (Kanno, 2000). • Participants described their multicultural experiences as stories along the lines of a “life narrative procedure”, with characters, settings, high and low points, and using chapters as an organizing framework (Yampolsky, Amiot & de la Sablonniere, 2013). 	<ul style="list-style-type: none"> • Measures are supported by following participants’ experiences over time, along with other observations. • This technique is appropriate for interpretive research designs. Transparency and appropriateness of this technique will vary across projects. 	<ul style="list-style-type: none"> • Narrative inquiry allows individuals to self-identify (i.e. identities are not imposed upon respondents) • Data collected by narrative inquiry allow for novel discoveries not anticipated by researchers in advance. • Data tend to be rich, such as observations about consistency or change over time, emotions, and sense making. 	<ul style="list-style-type: none"> • Narrative inquiry data are interpreted uniquely by each researcher, meaning this method is inappropriate for studies seeking consistency within the field.
Visual anthropology example: O’Connor (2011).	<ul style="list-style-type: none"> • Using visual anthropological methods, artifacts are examined as representations of multicultural identity. For example, Muslim Pakistani-Hong-Kong boys’ search for space to play cricket was interpreted through the lens of their hybrid cultural identities (O’Connor, 2011). 	<ul style="list-style-type: none"> • Interpretation of artifacts is justified differently within each paper. Reliability and validity are not the key criteria; instead, it is the quality of the interpretive story built by researchers. 	<ul style="list-style-type: none"> • Does not require individuals to have introspective access about their cultural identities. • Artifacts may be less susceptible to impression management than the content of interviews. • Data collected by these methods allow for novel discoveries not anticipated by researchers in advance. 	<ul style="list-style-type: none"> • Identities may be imposed on individuals through the researchers’ interpretation of artifacts.

Life Story Interview (Hammack, 2010).	<ul style="list-style-type: none"> ● “Draw a line that represents your life”. 	<ul style="list-style-type: none"> ● This technique is appropriate for constructivist research designs. Transparency and appropriateness of this technique will vary across projects. 	<ul style="list-style-type: none"> ● It allows individuals to self-identify (i.e. identities are not imposed upon respondents) ● Rich data are especially useful for examining changes to multicultural identification over time. 	<ul style="list-style-type: none"> ● Coding of open-ended responses can be difficult and time-consuming ● Range of interpretations of responses are possible, making it difficult to ensure participants’ intended meanings are represented authentically.
<i>Implicit measures</i>				
Implicit Association Test (IAT). Used by Devos (2006).	<ul style="list-style-type: none"> ● IATs assessing the strength of identification with the concepts “American culture” and “Mexican culture”, contrasted with “Other cultures” by measuring relative response times to pairing the word me (vs. them) with cultural words and pictures relating to Mexican culture and American culture. 	<ul style="list-style-type: none"> ● Devos (2006). Reliability: $\alpha = .66$ to $.89$ for identification with American culture stimuli; $.94$ to $.95$ for identification with Mexican culture stimuli; $.86$ to $.93$ for identification with other cultures stimuli; $.74$ to $.82$ for multi-item identification with American culture; $.67$ to $.88$ for multi-item identification with Mexican culture; $.58$ to $.90$ for multi-item identification with Asian culture. 	<ul style="list-style-type: none"> ● Measures implicit (i.e., subconscious) identity, which is a distinct construct from explicit identity. ● Does not require introspective access, although this is not generally a concern for reporting identification. ● Avoids self-presentational concerns (e.g., demand characteristics). ● Better able to predict behaviors than explicit measures. ● α for subscales with European American, Mexican American, and Asian American samples = $.58$ to $.95$ 	<ul style="list-style-type: none"> ● Stimuli must be carefully chosen and pre-tested. ● Predetermined cultural categories may not be appropriate (e.g., too broad, or not allow for hybrid culture). ● Forced comparison between Mexican and American cultures. ● More difficult to administer than scale measures; requires a lab setting. ● Assumes that cultural identities are meaning systems, so the distinction between identity, knowledge, and internalization is unclear. ● Validity information not reported.
Multiethnic Identities Processing Task (MIPT). Used by Marks, Patton, & Coll (2011).	<ul style="list-style-type: none"> ● Assess strength of ethnic identification through response times in a pairing task of various ethnic identity labels with “like me” versus “not like me”. 	<ul style="list-style-type: none"> ● Reliability and validity information not reported. 	<ul style="list-style-type: none"> ● Measures implicit (i.e., subconscious) identity, which is a distinct construct from explicit identity. ● Does not require introspective access, although this is not generally a concern for reporting identification. ● Avoids self-presentational concerns, where participants try 	<ul style="list-style-type: none"> ● Predetermined cultural categories may not be appropriate (e.g., too broad, may not allow for hybrid cultures). ● More difficult to administer than scale measures; requires a lab setting. ● No reliability or validity information were reported.

to identify themselves as they think the researchers want.

- Better ability to predict behaviors than explicit measures.
-

Supplementary Table 3. Review of existing measures of cultural internalization

◆ Indicates recommended measure

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
Acculturation scales				
<p>Vancouver Index of Acculturation (VIA) (Ryder, Alden, & Paulhus, 2000).</p> <p>Used by Kim & Hou (2016). Kim, Shen, Huang, Wang, & Orozco-Lapray (2014). David, Okazaki, Saw (2009).</p>	<ul style="list-style-type: none"> ● “I often behave in ways that are typical of my <i>heritage culture</i>”. ● “I believe in mainstream North American values”. 	<ul style="list-style-type: none"> ● Ryder et al. (2000). Reliability: $\alpha = .79$ to $.91$ for Heritage subscale; $\alpha = .75$ to $.89$ for Mainstream subscale. Mean inter-item correlations $r = .51$ to $.53$ for Heritage subscale and $r = .38$ to $.45$ for Mainstream subscale. Significant intercorrelations between Chinese and East Asian samples. Concurrent validity: Across Chinese, East Asian, and miscellaneous samples, Heritage subscale and Mainstream subscale had correlations of $r = .16$ to $.60$ with time lived in West, educated in West, generational status, anticipation of remaining in West, English as first language, and SL-Asia mean scores. Construct validity: 2 components extracted with Principal Components Analysis - heritage and mainstream identity; the 2 	<ul style="list-style-type: none"> ● Includes many aspects of life (e.g. social activities, entertainment). ● Includes behaviors plus belief in the values of each culture. ● Can be adapted to different cultures, including a hybrid culture (see Kim & Hou, 2016). ● Evidence for reliability and validity across multiple studies and samples. 	<ul style="list-style-type: none"> ● Scale is meant for two cultures only. ● “North American” (Ryder et al., 2000 scale) could be seen as more than one culture.

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
		<p>components were close to orthogonal ($r = .15$).</p> <ul style="list-style-type: none"> ● Kim, Shen, Huang, Wang, & Orozco-Lapray (2014). <p>Reliability: $\alpha = .83$ to $.88$ for Chinese (Heritage) subscale, and $.83$ to $.87$ for American (Mainstream) subscale.</p> <ul style="list-style-type: none"> ● Kim & Hou (2016). <p>Reliability: $\alpha = .86$ to $.89$ for Chinese (Heritage) subscale, $.79$ to $.88$ for American (Mainstream) subscale, and $.89$ to $.91$ for Chinese American orientation (new subscale with the term “Chinese American” used for the cultural identity).</p> <p>Validity: 1-factor confirmatory factor analysis on Chinese American subscale showed model fit; items loaded significantly on the factor across informants.</p> <ul style="list-style-type: none"> ● David, Okazaki, Saw (2009). <p>Reliability: $\alpha = .87$ for Heritage subscale, $\alpha = .86$ for Mainstream subscale.</p>		

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
<p>Revised Acculturation Rating Scale for Mexican Americans (ARSMA-II) (Cuéllar, Arnold & Maldonado, 1995).</p> <p>Used by Ferguson, Ferguson, & Ferguson (2017; adapted). Cuéllar, Nyberg, Maldonado, & Roberts (1997).</p>	<ul style="list-style-type: none"> • “I have difficulty accepting certain practices and customs commonly found in some Anglos”. • “I have difficulty accepting some values held by some Mexicans”. 	<ul style="list-style-type: none"> • Cuéllar, Arnold & Maldonado (1995). Reliability: $\alpha = .83$ for Anglo Orientation Subscale (AOS), $.88$ for Mexican Orientation Subscale (MOS), $.87$ for Marginal. Test-retest reliability with a one-week interval was $.94$ for AOS, $.96$ for MOS, $.78$ for Marginal. Concurrent validity: $r = .89$ between original ARSMA scale and ARSMA-II. Predictive validity: $r = .61$ between acculturation and generation status .ANOVA showed differences between generations. Construct validity: MOS showed 3 factors: language; ethnic identity, and ethnic interaction or ethnic distance. Anglo Orientation Subscale showed 2 factors: language and ethnic interaction or ethnic distance. • Cuéllar, Nyberg, Maldonado, & Roberts (1997). $\alpha = .83$ for Mexican Orientation Subscale (MOS), $\alpha = .88$ for Anglo Orientation Subscale (AOS). 	<ul style="list-style-type: none"> • Includes many facets of life (e.g. language, friends, and media). • Possibly could be adapted to different cultures. • Evidence reliability and validity. 	<ul style="list-style-type: none"> • Scale is meant for two cultures only. • Scale was developed for Mexican Americans; it may not be valid for other groups. • “Anglo,” may tap multiple cultures; it may fit ethnicity better than culture. • Accepting ideas from “some people” of a culture may not really capture internalization. • Includes items that relate to identification for MOS subscale only, conflating the two constructs. • Subscales have different items for each culture, reducing within-person comparability across cultures.
<p>Acculturation Index (AI) (Ward & Kennedy, 1994).</p>	<ul style="list-style-type: none"> • Participants answer the following questions about 	<ul style="list-style-type: none"> • Ward & Kennedy (1994). 	<ul style="list-style-type: none"> • Includes many facets of lifestyle. 	<ul style="list-style-type: none"> • Assume national-level cultural differences

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
Used by Tadmor, Galinsky, & Maddux (2012). Ward, Stuart, & Kus (2011). Tadmor, Tetlock & Peng (2009).	their lifestyle for 21 behavioral and cognitive items, (e.g., language, food, recreational activities, in-group and out-group perceptions): <ul style="list-style-type: none"> ○ “Are your experiences and behaviors similar to those of people from your country of origin (co-nationals)?” ○ “Are your experiences and behaviors similar to those of people from your host culture (host nationals)?” 	Reliability: $\alpha = .93$ for co-national internalization, $.96$ for host national internalization. Validity: co-national and host national internalization were independent $r = .23$. ● Tadmor, Galinsky, & Maddux (2012). Reliability: $\alpha = .59$ to $.74$ for home culture internalization, $\alpha = .71$ to $.77$ for host culture internalization, $\alpha = .91$ for Israeli internalization and $\alpha = .93$ for American internalization. Validity: Orthogonality of home and host culture internalization $r = .02$ to $.07$, and of Israeli and American culture internalization $r = .02$. Convergent validity: Significant relationships between AI and Benet-Martínez et al.’s (2006) American and culture of origin identification scale ($r = .57$ for American culture, $r = .46$ for Israeli culture). ● Tadmor, Tetlock, & Peng (2009). Reliability: $\alpha = .91$ to $.93$ for American internalization, $\alpha = .89$	<ul style="list-style-type: none"> ● Addresses psychological and sociocultural dimensions. ● Evidence of reliability and validity in assessing immigrants and sojourners with a range of cultural backgrounds (Ward, 1999; Ward & Rana-Deuba, 1999). ● Could be applied to other cultures due to culture-general wording. ● $\alpha = .59$ to $.96$ ● Evidence for convergent validity, independence, and orthogonality of dimensions. 	<ul style="list-style-type: none"> ● 21-item scale leaves less time to assess other constructs. ● Does not account for hybrid cultural internalization. ● Designed for sojourners; may not apply for all multiculturals (e.g., “country of origin” and “host culture” may not be appropriate terms). ● Dimensions are described in terms of cultural identification, although they primarily assess cultural internalization. One identity item is included along with the behavioral and cognitive items that assess internalization, conflating the two dimensions.

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
<p>Suinn-Lew Asian Self-Identity Acculturation Scale (SL-ASIA) (Suinn, Rickard-Figueroa, Lew, & Vigil, 1987).</p> <p>Used by Sirikantraporn, (2013).</p> <p>Dao, Teten, & Nguyen (2011).</p> <p>Chiao et al. (2010).</p>	<ul style="list-style-type: none"> • 21 multiple choice questions cover language, identity, friendship choice, behaviors, generation, geography, history, and attitudes. e.g., “What is your music preference? 1. Only Asian music (for example, Chinese, Japanese, Korean, Vietnamese, etc.) 2. Mostly Asian 3. Equally Asian and English 4. Mostly English 5. English only”. 	<p>for East Asian internalization. $\alpha = .91$</p> <p>for Israeli internalization.</p> <p>Validity: Orthogonality of Asian and American internalization scales $r = .00$ and Israeli and American internalization scales $r = -.02$.</p> <p>Convergent validity: Significant relationships between AI and Benet-Martínez et al.’s (2006) American and culture of origin identification scale: $r = .46$ to $.64$ for Asian, American, and Israeli cultures.</p> <ul style="list-style-type: none"> • Suinn, Rickard-Figueroa, Lew, & Vigil (1987). Reliability: $\alpha = .88$ Convergent validity: ANOVA showed significance by generation and length of residence in USA. ANOVA on 1-item scale of “how would you rate yourself?” on a scale ranging from “very Asian” to “very Anglicized” with “bicultural” in the middle was significant with means in the expected direction. • Dao, Teten, & Nguyen 	<ul style="list-style-type: none"> • Assesses multiple domains (e.g., language, behaviors) • Evidence of scale validity. • $\alpha = .79$ to $.91$. • Evidence for convergent validity. 	<ul style="list-style-type: none"> • Unidimensional scale (biculturalism = scalar midpoint score) implies one can only be strongly oriented toward one culture. • 21-item scale leaves less time to assess other constructs. • Specific to Asian and Western cultures; it may not be valid for other groups. • Conflates knowledge, demographics, and behaviors - does not measure internalization per se (e.g., of values). • Question format means this scale can only be used to assess internalization of two

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
		(2011). Reliability: $\alpha = .88$ total Asian sample, .89 Chinese sample, .88 Japanese sample, .91 Korean sample, .79 Filipino sample, .83 Vietnamese sample.		cultures, not more.
Acculturation strategies (Berry, Kim, Power, Young, & Bujaki, 1989). Used by Benet-Martínez, Lee, & Leu (2006).	<ul style="list-style-type: none"> ● 20-item measure of four acculturation strategies. <ul style="list-style-type: none"> ○ Separation, e.g., “I would rather marry a Chinese than an American”. ○ Assimilation, e.g., “I feel that the Chinese should adapt to American cultural traditions and not maintain their own”. ○ Integration, e.g., “It’s important to me to be fluent in both Chinese and English”. ○ Marginalization, e.g., “I prefer social activities that involve neither Americans nor Chinese”. 	<ul style="list-style-type: none"> ● Berry, Kim, Power, Young, & Bujaki (1989). Reliability: $\alpha = .68$ to .90 for the four acculturation strategies among French-, Portuguese-, Korean-, and Hungarian-Canadians. Validity: Face-validity: for the French-Canadian scale, bilingual judges sorted items into acculturation categories. Predictive validity: differences in acculturation strategy based on cultural club membership, newspaper readership, ethnic identification, and language participation. 	<ul style="list-style-type: none"> ● Assesses five domains (marriage, cultural traditions, language, social activities, and friends). ● Assesses both internalization of values and behavior. ● $\alpha = .68$ to .90 for the four acculturation strategies in four different cultural groups. ● Evidence for face and predictive validity across four different cultural groups. 	<ul style="list-style-type: none"> ● Forces participants into categories; does not treat internalization as a spectrum. ● 20-item scale leaves less time to assess other constructs. ● Specified cultural labels may not be appropriate - does not allow for a hybrid culture. ● Predictive validity varied somewhat based on sample. ● Question format means this scale can only be used to assess internalization of two cultures, not more.
Behavioral Acculturation Scale (Szapocznik, Scopetta, Kurtines, & Aranalde, 1978). Used by Szapocznik, Kurtines, & Fernandez (1980; adapted into Bicultural Involvement Questionnaire).	<ul style="list-style-type: none"> ● Participants report frequency with which they engage in each behavior - 24 items, e.g, “What sort of music do you listen to? 1. Cuban all the time 2. Cuban most of the time 3. Cuban at times and American other times 4. 	<ul style="list-style-type: none"> ● Szapocznik, Scopetta, Kurtines, & Aranalde (1978). Reliability: $\alpha = .97$ for the behavioral acculturation scale, .77 for the value acculturation scale. Parallel form reliability: Spanish and English 	<ul style="list-style-type: none"> ● Assesses multiple domains (e.g., language, music, celebrations). ● $\alpha = .77$ to .97. ● Evidence for parallel and test-retest reliability. ● Evidence for validity. 	<ul style="list-style-type: none"> ● Unidimensional scale (biculturalism score = Hispanicism score minus Americanism score, with scores close to zero indicating biculturalism) implies one must have relatively equivalent orientation with each culture

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
	<p>American most of the time 5. American all of the time”.</p>	<p>language scales had correlations of .88 ($p < .01$) for behavioral acculturation and .46 ($p < .01$) for value acculturation.</p> <p>Test-retest reliability: test-retest correlation of both scales with a new sample was .96 for behavioral acculturation and .86 for value acculturation ($p < .01$ for both).</p> <p>Validity: Factor analysis: 4 factors accounted for 28.7% of the total variance. Factor loadings .37 or above for each item.</p> <p>Discriminant validity: Significance in item discrimination for comparison between item-by-item responses of the Cuban respondents and the cultural referent group.</p> <p>Index of acculturation discrimination created through comparing item-by-item responses of highly and low acculturated Cubans.</p> <p>Criterion-related validity: Length of time in host country correlated with behavioral acculturation (.61 $p < .01$) and values acculturation (ranged from .31 to .38, $p < 0.01$). Inter-</p>		<p>to be bicultural.</p> <ul style="list-style-type: none"> ● 24-item scale leaves less time to assess other constructs. ● Specific to Cuban or Hispanic and American cultures; it may not be valid for other groups. ● Measures behaviors, preferences, and enjoyment of activities rather than internalized values. ● Question format means this scale can only be used to assess internalization of two cultures, not more, and not internalization of hybrid cultures.

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
		generational differences significantly related to behavioral acculturation ($F(4, 319); 47.01, p < .01$).		
Cultural adaptation (Hanek, Lee, & Brannen, 2014; adapted from Lin & Malhotra, 2012).	<ul style="list-style-type: none"> ● “To what extent did you adapt your thoughts to this culture?” 	<ul style="list-style-type: none"> ● Hanek, Lee & Brannen (2014). Reliability: $\alpha = .82$. 	<ul style="list-style-type: none"> ● Assesses adaptation of thoughts and behaviors, as well as knowledge. ● Could be applied to other cultures due to culture-general phrasing. 	<ul style="list-style-type: none"> ● Conflates internalization, knowledge and behaviors. ● No validity information provided.
Cultural identity (Hanek, Lee, & Brannen, 2014).	3-item scale: <ul style="list-style-type: none"> ● “To what extent do you identify with this country?” ● “To what extent are your values the same as this country?” ● “To what extent is your worldview the same as this country?” 	<ul style="list-style-type: none"> ● Hanek, Lee & Brannen (2014). Reliability: $\alpha = .73$ for identification with home culture, $\alpha = .80$ for identification with host culture. No validity information reported. 	<ul style="list-style-type: none"> ● 3-item scale is simple to administer. ● Could be applied to other cultures due to culture-general phrasing. 	<ul style="list-style-type: none"> ● Purportedly a measure of cultural identity, but two out of three scale items are about internalization – it conflates identity and internalization. ● Use of the term “country” implies homogenous national-level culture. ● No validity information provided.
Cultural Life Style Inventory (CLSI) (Mendoza, 1989). Used by Luna, Ringberg, & Peracchio (2008).	<ul style="list-style-type: none"> ● Scale ranging from “only Hispanic” to “only Anglo” with the midpoint “Hispanic and Anglo about equally”, e.g., “Ethnicity of friendship ties”; “Culture subject feels most comfortable with”; “Ethnic foods subject eats”. ● Subscales include intrafamilial language use, extrafamilial language use, social affiliation and activities, cultural familiarity and activities, and cultural identification and pride. 	<ul style="list-style-type: none"> ● Mendoza (1989). Reliability: $\alpha = .84$ to .91 for subscales. Test-retest reliability: correlations ranged from $r = .88$ to .95 for Mexican-Americans taking it in English, Mexican-Americans taking it in Spanish, and Anglo-Americans. Parallel forms: Spanish and English version correlations of $r = .80$ for English and then Spanish versions being taken, and $r = .77$ for the opposite. Face validity: Anglo- 	<ul style="list-style-type: none"> ● Full scale also includes identity and language usage subscales. ● $\alpha = .84$ to .91. ● Evidence for multiple types of validity. 	<ul style="list-style-type: none"> ● Assesses behavior and attitude rather than internalization. ● Cannot assess level of internalization for both cultures. The scalar midpoint suggests biculturalism, and endpoints monoculturalism. ● Question format means this scale can only be used to assess internalization of two cultures, not more, and not the degree of internalization of a hybrid culture.

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
		<p>American and Mexican-American judges rated the extent to which items tapped into acculturation.</p> <p>Content validity: Items with the highest mean ratings by judges were tested for item discrimination; only retained items that had respondents typical of 90% or more of Mexican-American respondents and 90% or more of Anglo-American respondents.</p> <p>Construct validity: Principal components, cluster analysis, and multidimensional scaling performed. All methods showed 5 distinct factors.</p> <p>Generational level associated with reduced Mexican culture and increased Anglo-American culture at a monotonic rate ($\eta = .54$ for Mexican, $.61$ for Anglo-American.).</p> <p>Exposure to mainstream culture positively related to cultural shift ($r = .66, p < .001$) and negatively related to cultural resistance and incorporation ($r = -.60$, and $r = -.25$, respectively).</p> <p>Predictive validity: Transient and temporary immigration status related to more</p>		

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
		cultural resistance ($\eta = .46$), less cultural shift ($\eta = .42$), and less cultural incorporation ($\eta = .39$). Concordance: significant correlation between self-report and immediate family's report about respondents ($r = .71$).		
Mexican American Biculturalism Scale (MABS) (Basilio, Knight, O'Donnell, Roosa, Gonzales, Umaña-Taylor, & Torres, 2014).	<ul style="list-style-type: none"> • Bicultural facility items, e.g., “Needing to speak Spanish sometimes, and English other times is ...” (very easy to very difficult). • 3 factors: bicultural comfort, bicultural facility, and bicultural advantages. 	<ul style="list-style-type: none"> • Basilio, Knight, O'Donnell, Roosa, Gonzales, Umaña-Taylor, & Torres (2014). Reliability: $\alpha = .81$ to $.92$ for subscales of bicultural comfort, bicultural facility, bicultural advantages, and overall biculturalism. Median inter-item correlations ranged from $.33$ to $.52$ for all subscales. <p>Construct validity: Confirmatory factor analysis showed 3 factors, consistent across age groups, language, and gender. English use, Spanish use, language conflict, English pressures, Spanish pressures, perceived discrimination, and ethnic identity significantly related to the factors and the overall scale.</p>	<ul style="list-style-type: none"> • $\alpha = .81$ to $.92$ for all subscales. • Evidence of scale validity. • Emic approach; scale was designed for a specific population, so is tailored to that group. 	<ul style="list-style-type: none"> • Assesses attitudes and competency rather than internalization. • Developed for Mexican-American sample; may not be valid for other cultures. • Question format means this scale can only be used to assess internalization of two cultures, not more.
Open-ended questions: Used by van de Vijver,	<ul style="list-style-type: none"> • “Where do your three closest friends come from?” 	<ul style="list-style-type: none"> • van de Vijver, Blommaert, Gkoumasi, & Stogianni (2015). 	<ul style="list-style-type: none"> • Allows participants to describe internalization in their own words. 	<ul style="list-style-type: none"> • Some questions rely on introspective access by participants, which

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
Blommaert, Gkoumasi, & Stogianni (2015).	<ul style="list-style-type: none"> “Could you describe a number of aspects that you really find important in life?” 	<p>Validity: Factor analysis including three open-ended questions, answers to 10 statements of “who am I?”, and five identity total scores (ethnic, Belgian, familial, religious, and cosmopolitan) suggested two factors: identity and belongingness.</p>	<ul style="list-style-type: none"> Open-ended questions may lead to rich insights not anticipated in advance by the researchers. Evidence for validity when combined with other responses. 	<p>participants may find challenging.</p> <ul style="list-style-type: none"> Some questions may be assessing behavior rather than internalization. Some questions may be conflating identity with internalization.
Behavior				
<p>Individuating behavior (Maslach, 1974).</p> <p>Used by Chen, Lam, Hui, Ng, Mak, Guan, Buchtel, Tang, & Lau, (2016).</p> <p>Chen, Bond, Chan, Tang, & Buchtel, (2009).</p>	<ul style="list-style-type: none"> 12 item scale assesses the willingness to differentiate oneself publicly, e.g., (To what extent do you) “Give your opinion on a controversial issue, even though no one has asked for it”. 2 factors: individuation-singular and individuation-personal. 	<ul style="list-style-type: none"> Maslach (1974). <p>Validity: Factor analysis of behaviors that people in the individuating condition engaged in suggested 2 factors.</p> <ul style="list-style-type: none"> Chen, Lam, Hui, Ng, Mak, Guan, Buchtel, Tang, & Lau (2016). <p>Reliability: $\alpha = .87$.</p> <ul style="list-style-type: none"> Chen, Bond, Chan, Tang, & Buchtel (2009). <p>Reliability: $\alpha = .84$.</p>	<ul style="list-style-type: none"> Cross-cultural differences have been found on individuating behavior. $\alpha = .84$ to $.87$ Evidence for validity of 2 factors. 	<ul style="list-style-type: none"> Only assesses one aspect of behavior. Behavior is not internalization. Theoretical underpinning for using this scale to assess multicultural internalization is not explained.
<p>Modest behavior scale (MBS) (Chen, Bond, Chan, Tang, & Buchtel, 2009).</p> <p>Used by Chen, Lam, Hui, Ng, Mak, Guan, Buchtel, Tang, & Lau, (2016).</p>	<ul style="list-style-type: none"> (To what extent do you) “Deny my own strengths in front of others.” Used to assess multiculturalism through changing behaviors 	<ul style="list-style-type: none"> Chen, Bond, Chan, Tang, & Buchtel (2009). <p>Reliability: $\alpha = .38$ to $.80$ for self-effacement, other-enhancement, and avoidance of attention-seeking behavior.</p> <p>Construct validity: EFAs and Cattell’s scree tests in three studies suggest 3 factors, each accounting for between 7% and 16% of variance.</p>	<ul style="list-style-type: none"> Cross-cultural differences have been found on modest behavior. Other variables were mostly related as expected to MBS sub-factors, but not entirely consistently. 	<ul style="list-style-type: none"> 39-item scale leaves less time to assess other constructs. Only assesses one aspect of behavior. Behavior is not internalization.

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
		<p>Predictive validity: Different sub-factors significantly related to trait modesty, traditionalism, self-efficacy, individuation, conservation, independence, interdependence, self-enhancement, and openness to change.</p> <ul style="list-style-type: none"> • Chen, Lam, Hui, Ng, Mak, Guan, Buchtel, Tang, & Lau (2016). <p>Reliability: $\alpha = .83$.</p>		
Values				
<p>Family Relationship Values (Berry, Phinney, Sam, & Vedder, 2006).</p> <p>Used by Ferguson, Ferguson, & Ferguson (2017; used the family obligations sub-scale only).</p>	<ul style="list-style-type: none"> • Two subscales: <ul style="list-style-type: none"> ○ Family obligations (10 items), e.g., “Children should obey their parents”. ○ Adolescent rights (4 items), e.g., “When a girl reaches the age of 16, it is all right for her to decide whom to date”. 	<ul style="list-style-type: none"> • Berry et al. (2006). <p>Predictive validity: indicated through integration acculturation profile (i.e., multicultural individuals) being associated with acceptance of both obligations and rights within their families.</p> <p>Reliability: $\alpha = .72$ for family obligations subscale (immigrants and nationals); $\alpha = .78$ (immigrants) and $.75$ (nationals) for adolescents’ rights subscale. Authors found “very strong structural support for the structural equivalence of the measures” (p.312) across cultural groups and</p>	<ul style="list-style-type: none"> • Culture-general. • Predictive validity of both subscales indicated through multicultural (i.e., integration) acculturation profile being associated with acceptance of both obligations and rights within families (Berry et al., 2006). • Predictive validity - two clusters show significantly different values for Family obligations and interdependence respectively ((M)ANOVA test) (Ferguson et al., 2017). 	<ul style="list-style-type: none"> • Theoretical underpinning for using this scale to assess multicultural internalization is not explained. • Limited evidence for validity of adolescents’ rights subscale. • Evidence only of predictive validity (but not other types of validity) for family obligations subscale. • Some items may not fit all contexts, e.g., asking about dating will not work in cultures where dating is uncommon.

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
		<p>countries.</p> <ul style="list-style-type: none"> ● Ferguson et al. (2017). <p>Predictive validity: indicated through cluster analysis showing lower family obligations in the same cluster as lower remote culture orientation and less interdependent self-construal, in line with theory.</p> <p>Reliability: $\alpha = .65$ for family obligations subscale.</p>		
<i>Self-construals</i>				
<p>Bem Sex Role Inventory (BSRI) (Bem 1979).</p> <p>Used by Luna, Ringberg, & Peracchio (2008; reduced form).</p>	<ul style="list-style-type: none"> ● 20-item scale. Participants rate themselves on feminine and masculine traits, which Luna et al. (2008) characterize as self-sufficiency versus other-dependence self-construals. 	<ul style="list-style-type: none"> ● Bem (1979). Reliability and validity information not reported. ● Luna et al. (2008). Convergent validity: Analysis of Bem's scale supported qualitative findings that the Spanish language cued self-sufficient self; English language cued the other-dependent self. <p>Other validity information not reported.</p>	<ul style="list-style-type: none"> ● Emic approach (measure adapted specifically for Hispanic female sample), so internalization assessment is tailored to this group. ● Convergent validity indicated in Luna, Ringberg, and Peracchio (2008) by supportive qualitative findings. 	<ul style="list-style-type: none"> ● Scale is designed to measure masculinity-femininity and gender roles; it was not designed or validated as a measure of cultural internalization. ● No information on reduced form scale reliability (as a measure of cultural internalization among Hispanic women). ● Specific to Hispanic women; may not be valid for other cultures. ● Scale reliabilities not reported.
<p>Self-Construal Scale (SCS) (Singelis, 1994).</p> <p>Used by Yamada & Singelis (1999).</p>	<ul style="list-style-type: none"> ● Independent subscale (12 questions), e.g., "My personal identity, independent of others is important to me". ● Interdependent subscale 	<ul style="list-style-type: none"> ● Singelis (1994). Construct validity: Asian Americans were more interdependent than Caucasian Americans. 	<ul style="list-style-type: none"> ● Could be applied to other cultures due to culture-general phrasing. ● Evidence for reliability and validity. ● Theoretical basis for scale 	<ul style="list-style-type: none"> ● 26-item scale leaves less time to assess other constructs. ● Only assesses one aspect of internalization (independent versus interdependent self-

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
	(14 questions), e.g., “It is important for me to maintain harmony within my group”.	were more independent than Asian Americans. The two subscales were found to be orthogonal using factor analysis across two samples. Divergent validity: Independence scores varied between ethnic groups, but were not associated with attributions to the situation. Predictive validity: Independent subscale predicted attribution better than ethnic group. Interdependent subscale accounted for variance in attributions after controlling for ethnic group. Reliability: $\alpha = .73$ (sample 1) and $.74$ (sample 2) for interdependent subscale; $\alpha = .69$ (sample 1) and $.70$ (sample 2) for independent subscale. ● Yamada & Singelis, (1999). Validity information not reported. Reliability: $\alpha = .74$ for interdependent subscale; $.67$ for independent subscale.	dimensions.	construals).
Individual- and Social-Oriented Self (Lu, 2007).	Two subscales: ● 20 items measuring	● Lu (2008). Factor structure: CFA	● Theoretical basis for scale dimensions.	● 40-item scale leaves less time to assess other

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
Used by Lu (2008).	<p>individual-oriented self (ISS-I), comprising four elements: independence, self-determination, competition, consistency.</p> <ul style="list-style-type: none"> ● 20 items measuring social-oriented self (ISS-S), comprising four elements: contextual self, interpersonal relatedness, self-cultivation, social sensitivity. 	<p>supported the 2-factor structure of individual-oriented self (ISS-I) and social-oriented self (ISS-S).</p> <p>Convergent validity: Strong, positive correlations between: ISS-I and independent self-construal scale; and ISS-S and interdependent self-construal scale. Positive correlations between ISS-S and both horizontal and vertical collectivism, but ISS-I failed to correlate with horizontal and vertical individualism.</p> <p>Divergent validity: No or weak correlations between ISS-I and interdependence; and ISS-S and independence.</p> <p>Predictive validity: ISS-I generally showed correlations as expected with a range of measures, such as ego-focused positive emotions, and individual-oriented achievement motivation. ISS-S was generally correlated as expected with a range of measures, such as communal orientation, harmony beliefs, holistic thinking,</p>	<ul style="list-style-type: none"> ● Useful for “contemporary Chinese individuals” - designed to capture unique aspects of Chinese self. ● Reliability and validity was demonstrated across five studies in Lu (2008). Evidence provided for factor structure, convergent validity, divergent validity, predictive validity, and reliabilities. 	<p>constructs.</p> <ul style="list-style-type: none"> ● Specific to Chinese culture; may not be valid for other cultures. ● Only measures one aspect of internalization – individual- and social-oriented self-construals.

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
		and social-oriented achievement motivation. Reliability: $\alpha = .79$ for ISS-I subscale; $\alpha = .84$ for ISS-S subscale. Test-retest reliability $r = .70$ for ISS-I and $r = .57$ for ISS-S.		
<p>◆ Self-Construal Scale (Gudykunst, Matsumoto, Ting-Toomey, Nishida, Kim, & Heyman, 1996).</p> <p>Used by Chen et al., (2016).</p>	<ul style="list-style-type: none"> Independent (14 items) and interdependent (15 items) views of the self, e.g., I try not to depend on others [independent]; I consult with others before making important decisions [interdependent]. 	<ul style="list-style-type: none"> Gudykunst et al. (1996). Construct validity: “Given the theoretical rationale, the analysis was restricted to a two-factor solution... Twenty-nine of the items loaded on one of the two factors” (p. 526). <p>Predictive validity: scale shown to account for more variance in low context and high context communication styles than does individualism-collectivism measures.</p> <p>Reliabilities: $\alpha = .80$ to $.85$ for interdependent self-construal across four samples (United States, Japan, Korea, and Australia); $\alpha = .73$ to $.83$ for independent self-construal across four samples.</p> <ul style="list-style-type: none"> Chen et al. (2016). Convergent validity: multicultural acquisition was positively related to both independent and interdependent self- 	<ul style="list-style-type: none"> Items identified from an analysis across five cultural groups. Theoretically based scale dimensions. Evidence for scale reliability and validity. 	<ul style="list-style-type: none"> 20-item scale leaves less space to assess other constructs. Only measures one aspect of internalization.

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
Chinese self and Western self (Ng, 2007). Used by Ng, Ng, & Ye (2016).	<ul style="list-style-type: none"> Rate on a two-item scale, 1 = weak, 7 = strong, e.g., “I feel the Chinese identity in me is...”; “I feel the influence of Chinese culture on me is...” 	<p>construal. Reliabilities: $\alpha = .68$ for independence; $\alpha = .73$ for interdependence.</p> <ul style="list-style-type: none"> Ng, Ng, & Ye (2016), referring to findings originally reported in Ng, Rochelle, Shardlow, & Ng (2014) and Ng, Yam & Lai (2007). <p>Convergent validity: “Chinese self was positively related to proficiency and use of Chinese language and Chinese cultural practices..., whereas Western self was positively related to proficiency and use of English language and Western cultural practices.” (p.542). Construct validity: Cultural selves were “confirmed in a factor analysis on a sample of British Chinese immigrants from Hong Kong (Ng, et al., 2014)” (p. 542).</p> <ul style="list-style-type: none"> Ng, Ng, & Ye (2016). Reliability: $\alpha = .75$ for Chinese self subscale; $\alpha = .64$ for Western self subscale. 	<ul style="list-style-type: none"> Simple to administer. $\alpha = .64$ (Chinese items) to .75 (Western items). Evidence for validity. 	<ul style="list-style-type: none"> Conflates identification and internalization: the first item assesses identification, while the second item assesses internalization. Single dimension is inadequate to measure identification and internalization. Relies on introspective access to report internalization, which participants may find challenging. Uses the midpoint as a cutoff for strong self, which treats cultural self as a dichotomy. Reliability cannot be adequately assessed with only two items. Specific to Chinese and Western cultures; may not be valid for other cultures.
Implicit measures				
<ul style="list-style-type: none"> Spontaneous inferences from cultural cues (measure of cultural values) (Fu, Chiu, 	<ul style="list-style-type: none"> Participants are asked to judge whether a given probe word appeared in the 	<ul style="list-style-type: none"> Fu et al. (2007). Reliability indicated by consistent results across 	<ul style="list-style-type: none"> Does not require introspective access to assess internalization, 	<ul style="list-style-type: none"> Task needs to be carried out in a laboratory setting. Requires careful design

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
Morris, & Young, 2007).	preceding sentence.	two studies (i.e., with two samples – Hong Kong undergraduates and Chinese Americans). In experimental trials: “The percentage of correctly made “no” responses in the experimental trials was high (98.4%), and the percentage of correctly made “yes” responses in the filler trials was lower (84.1%)” (p.63).	<p>meaning participants do not need to be aware of their own levels of cultural internalization.</p> <ul style="list-style-type: none"> ● Disguised as a decision-making task, so avoids self-presentational concerns, where participants try to respond consistently with researchers’ expectations. ● Taps into moral values, but could be modified to tap into other aspects of internalization (e.g., beliefs). ● Measures frame switching with a within-subject design, and can capture repeated instances of frame switching in a single session. ● Could be adapted to measure presence of hybrid cultural schemas. ● Taps into insider cultural knowledge rather than superficial knowledge. ● Reliability indicated through consistent findings across two studies (Fu et al., 2007). 	<p>(choice of suitable sentences and probe words for each culture of interest).</p> <ul style="list-style-type: none"> ● Relies on participants having deep “insider knowledge” in order for the spontaneous inferences to occur, so the measure may conflate knowledge and internalization. ● Has not been tested with a range of multiculturals – only Chinese and American cultures. ● Validity information not reported.
Frame switching on cultural values in a reaction time task (Chao, Chen, Roisman, & Hong, 2007).	<ul style="list-style-type: none"> ● Participants are asked to identify whether each string of letters was a real word. Some strings were culture-specific target words (e.g., <i>obedient</i>). Response times indicate whether or not participants accessed cultural schemas while processing their answers. 	<ul style="list-style-type: none"> ● Reliability and validity information not reported. 	<ul style="list-style-type: none"> ● Does not require introspective access to assess internalization, meaning participants do not need to be aware of their own levels of cultural internalization. ● Disguised as a perceptual performance task, so avoids self-presentational concerns 	<ul style="list-style-type: none"> ● Task needs to be carried out in a laboratory setting. ● Target words need to be carefully chosen for each culture of interest. ● Requires specialist software (DirectRT). ● Requires participants to undergo practice trials. ● Task not designed to directly

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
			(e.g., demand characteristics). <ul style="list-style-type: none"> ● Potential measure of frame switching with a within-subject design may capture repeated instances of frame switching in a single session. 	measure presence of cultural schema, but rather how a second variable (racial essentialism) facilitates or inhibits frame switching – unclear what the threshold would be to determine presence or absence of cultural schema. <ul style="list-style-type: none"> ● Validity and reliability not reported.
Frame switching on cooperation in prisoner's dilemma game and commons dilemma game (Brannon, Markus, & Taylor, 2015).	<ul style="list-style-type: none"> ● Measures participant changes in response to cultural primes, while engaging in cooperation games. 	<ul style="list-style-type: none"> ● Brannon et al.(2015). Validity information not reported. ● In pilot experiment, interrater reliability of coding of descriptions, mean K = .89. ● In studies 1 to 5, reliabilities not reported. 	<ul style="list-style-type: none"> ● Does not require introspective access to assess internalization, meaning participants do not need to be aware of their own levels of cultural internalization. ● Priming is not linked to national cultures, and therefore can be used to investigate research questions involving sub-national cultures (e.g., African-American). ● Evidence of reliability. 	<ul style="list-style-type: none"> ● Task needs to be carried out in a laboratory setting. ● Measures only one aspect of internalization (self-concept). ● Interrater reliabilities for Studies 1 to 5 were not reported. ● Validity not reported.
<ul style="list-style-type: none"> ◆ Implicit association test (IAT) (measure of frame switching on self-concept) (Greenwald, McGhee, & Schwartz 1998). <p>Used by Luna, Ringberg, & Peracchio (2008).</p>	<ul style="list-style-type: none"> ● Measures the relative automatic association between “self-sufficient” and “masculine” versus “self-sufficient” and “feminine”. 	<ul style="list-style-type: none"> ● Greenwald, McGhee, & Schwartz (1998). Experiment 1 - Divergent and convergent validity: Correlations between explicit measures of different contrasts (average $r = .41$) and between implicit measures of different contrasts (average $r = .58$) were much higher than the 	<ul style="list-style-type: none"> ● Does not require introspective access to assess internalization, meaning participants do not need to be aware of their own levels of cultural internalization. ● Avoids self-presentational concerns, where participants try to respond consistently with researcher expectations. ● Evidence of validity. 	<ul style="list-style-type: none"> ● Study design is specific to research question and sample, and would need to be adapted for other research questions and cultures; stimulus words need to be carefully chosen for each culture of interest. ● Measures only one aspect of internalization (self-concept). ● Reliability not reported.

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
		<p>correlations between explicit and implicit measures of the same contrast (average $r = .19$).</p> <p>Experiment 2 –</p> <p>Divergent validity: IAT was found to differentiate better between the Korean and Japanese subsamples when immersion in Asian culture was higher, as revealed in statistically significant difference in slopes for the subsample regression functions.</p> <ul style="list-style-type: none"> • Luna et al. (2008). <p>Not reported.</p>	<ul style="list-style-type: none"> • Emic approach (stimuli are tailored specifically for each sample). • Could be adapted for other cultures. 	<ul style="list-style-type: none"> • Stimuli must be created and validated for the cultures in each sample.
<p>Open-ended interpretation of stimuli (measure of self-concept).</p> <p>Used by Luna, Ringberg, & Peracchio (2008). Ng & Houston (2006).</p>	<ul style="list-style-type: none"> • Participants asked to interpret stimuli, e.g., “Please tell me your thoughts and feelings about this advertisement”. 	<ul style="list-style-type: none"> • Luna, Ringberg, & Peracchio (2008). Reliability: Analysis of the Bem scale supported the qualitative findings. • Ng & Houston (2006). Reliability: Inter-rater reliability = .95. <p>Convergent validity: Measure can distinguish between participants’ countries and supports qualitative findings</p>	<ul style="list-style-type: none"> • Can be used to assess several aspects of internalization (values, beliefs, etc.). • Evidence of reliability and convergent validity. 	<ul style="list-style-type: none"> • Coding is time-consuming. • Administering is time-consuming and labor-intensive (one-on-one interviews). • Responses may reflect variation in language ability rather than underlying differences in self-construal.
Interviews				
<p>Narrative inquiry</p> <p>Used by Kanno (2000).</p>	<ul style="list-style-type: none"> • Personal stories that emerge from interviews to reveal a bicultural self. 	<ul style="list-style-type: none"> • Transparency and appropriateness of this technique will vary across projects. 	<ul style="list-style-type: none"> • Rich data where reliability is supported by reporting participants’ own words wherever possible. • This technique is most 	<ul style="list-style-type: none"> • Time consuming. Interview design requires introspective access on internalization by participants, which is not

Internalization measure and usage	Sample items or techniques	Reliability/validity	Pros	Cons
			appropriate for interpretive research designs.	always achieved. E.g., Oswald (1999) reports discrepancies between Haitian-American participants' stated identities and behavior observed during the ethnography.
<p>Specific questions</p> <p>Used by Barker (2015).</p>	<ul style="list-style-type: none"> • “How have your ways of thinking, feeling, and acting changed during your time in Sweden?” 	<ul style="list-style-type: none"> • Barker (2015). Validity: Author claimed that validity was ensured in several ways, including counting occurrences and cases related to each major theme; analysing all disconfirming evidence, outliers, and exceptions; and “member checking” (respondent validation) of findings and interpretations. Reliability: Author stated that sampling, data collection, and analysis processes were transparent and “data were...scrutinized for possible interviewer effects” (p. 60). 	<ul style="list-style-type: none"> • Allows participants to describe, which can provide more nuanced insights than quantitative measures. • Reliability and validity was reportedly achieved in various ways. 	<ul style="list-style-type: none"> • Time consuming. • Requires introspective access into internalization by participants, meaning participants must be aware of their own levels of internalization. Some may find this difficult to report.

References for Supplemental Tables

- Barker, G. 2017. Acculturation and bicultural integration in organizations: Conditions, contexts, and challenges. *International Journal of Cross Cultural Management*, 17(3): 281-304.
- Barker, G. 2015. Choosing the best of both worlds: The acculturation process revisited. *International Journal of Intercultural Relations*, 45:56-69.
- Basilio, C. D., Knight, G. P., O'Donnell, M., Roosa, M. W., Gonzales, N. A., Umaña-Taylor, A. J., & Torres, M. 2014. The Mexican American Biculturalism Scale: Bicultural comfort, facility, and advantages for adolescents and adults. *Psychological Assessment*, 26(2): 539-554.
- Bem, S. L. 1979. Theory and measurement of androgyny: A reply to the Pedhazur-Tetenbaum and Locksley-Colten critiques. *Journal of Personality and Social Psychology*, 37(6): 1047-1054.
- Benbow, A., & Rutland, A. 2017. Competence matters! Understanding biculturalism in ethnically diverse adolescents. *Journal of Community & Applied Social Psychology*, 27(5): 366-373.
- Benet-Martínez, V., & Haritatos, J. 2005. Bicultural identity integration (BII): Components and psychosocial antecedents. *Journal of Personality*, 73(4): 1015-1049.
- Benet-Martínez, V. & Karakitapoğlu-Aygün, Z. 2003. The interplay of cultural values and personality in predicting life-satisfaction: Comparing Asian Americans and European Americans. *Journal of Cross-Cultural Psychology*, 34(1): 38-60.
- Benet-Martínez, V., Lee, F., & Leu, J. 2006. Biculturalism and cognitive complexity: Expertise in cultural representations. *Journal of Cross-Cultural Psychology*, 37(4): 386-407.
- Benet-Martínez, V., Leu, J., Lee, F., & Morris, M. W. 2002. Negotiating biculturalism: Cultural frame switching in biculturals with oppositional versus compatible cultural identities. *Journal of Cross-Cultural Psychology*, 33(5): 492-516.
- Berry, J., Kim, U., Power, S., Young, M., & Bujaki, M. 1989. Acculturation attitudes in plural societies. *Applied Psychology: An International Review*, 38(2): 185-206.
- Berry, J., Phinney, J. S., Sam, D. L., & Vedder, P. 2006. Immigrant youth: Acculturation, identity, and adaptation. *Applied Psychology*, 55(3): 303-332.
- Brannon, T. N., Markus, H. R., & Taylor, V. J. 2015. "Two souls, two thoughts," two self-schemas: Double consciousness can have positive academic consequences for African Americans. *Journal of Personality & Social Psychology*, 108(4): 586-609.
- Cameron, J. E. 2004. A three-factor model of social identity. *Self & Identity*, 3(3): 239-262.
- Carpentier, J., & de la Sablonnière, R. 2013. Identity profiles and well-being of multicultural immigrants: The case of Canadian immigrants living in Quebec. *Frontiers in Psychology*, 4(80): 1-50.
- Carrera, S. G., & Wei, M. 2014. Bicultural competence, acculturative family distancing, and future depression in Latino/a college students: A moderated mediation model. *Journal of Counseling Psychology*, 61(3): 427-436.
- Chao, M. M., Chen, J., Roisman, G. I., & Hong, Y.-y. 2007. Essentializing race: Implications for bicultural individuals' cognition and physiological reactivity. *Psychological Science*, 18(4): 341-348.
- Chen, S. X., Benet-Martínez, V., & Bond, M. H. 2008. Bicultural identity, bilingualism, and psychological adjustment in multicultural societies: Immigration-based and globalization-based acculturation. *Journal of Personality*, 76(4): 803-838.
- Chen, S.X., Bond, M. H., Chan, B., Tang, D., & Buchtel, E. E. 2009. Behavioral manifestations of modesty. *Journal of Cross-Cultural Psychology*, 40(4): 603-626.
- Chen, S. X., Lam, B. P., Hui, B. H., Ng, J. K., Mak, W. S., Guan, Y., Buchtel, E., Tang, W., & Lau, V. Y. 2016. Conceptualizing psychological processes in response to globalization: Components, antecedents, and consequences of global orientations. *Journal of Personality and Social Psychology*, 110(2): 302-331.
- Cheng, C., Lee, F., & Benet-Martínez, V. 2006. Assimilation and contrast effects in cultural frame switching: Bicultural identity integration and valence of cultural cues. *Journal of Cross-Cultural Psychology*, 37(6): 742-760.
- Chiao, J. Y., Harada, T., Komeda, H., Li, Z., Mano, Y., Saito, D., Parrish, T., Sadato, N., & Iidaka, T. 2010. Dynamic cultural influences on neural representations of the self. *Journal of Cognitive Neuroscience*, 22(1): 1-11.
- Chung, R. H. G., Kim, B. S. K., & Abreu, J. M. 2004. Asian American multidimensional acculturation scale: Development, factor analysis, reliability, and validity. *Cultural Diversity and Ethnic Minority Psychology*, 10(1): 66-80.
- Cross, S. N., & Gilly, M. C. 2014. Cultural competence and cultural compensatory mechanisms in binational households. *Journal of Marketing*, 78(3): 121-139.
- Cuéllar, I., Arnold, B., & Maldonado, R. 1995. Acculturation rating scale for Mexican Americans-II: A revision of the original ARSMA Scale. *Hispanic Journal of Behavioral Sciences*, 17(3): 275-304.

- Cuéllar, I., Nyberg, B., Maldonado, R., & Roberts, R. 1997. Ethnic identity and acculturation in young adult Mexican-origin population. *Journal of Community Psychology*, 25(6): 535-549.
- Dao, T. K., Teten, A., & Nguyen, Q. 2011. Linear and orthogonal models of acculturation and its relations to cultural variables: An examination of the Suinn-Lew Asian Self-Identity Acculturation Scale (SL-ASIA). *International Journal of Intercultural Relations*, 35(1): 61-68.
- David, E. J. R., Okazaki, S., & Saw, A. 2009. Bicultural self-efficacy among college students: Initial scale development and mental health correlates. *Journal of Counseling Psychology*, 56(2): 211-226.
- Der-Karabetian, A., & Ruiz, Y. 1997. Affective bicultural and global-human identity scales for Mexican-American adolescents. *Psychological Reports*, 80(3): 1027-1039.
- Devos, T. 2006. Implicit bicultural identity among Mexican American and Asian American college students. *Cultural Diversity and Ethnic Minority Psychology*, 12(3): 381-402.
- Dey, B., Balmer, J., Pandit, A., Saren, M., & Binsardi, B. 2017. A quadripartite approach to analysing young British South Asian adults' dual cultural identity. *Journal of Marketing Management*, 33(9-10): 789-816.
- Dimitrova, R., Chasiotis, A., Bender, M., & van de Vijver, F. 2013. Collective identity and wellbeing of Roma minority adolescents in Bulgaria. *International Journal of Psychology*, 48(4): 502-513.
- Downie, M., Mageau, G. A., Koestner, R., & Liodden, T. 2006. On the risk of being a cultural chameleon: Variations in collective self-esteem across social interactions. *Cultural Diversity and Ethnic Minority Psychology*, 12(3): 527-540.
- Ferguson, Y. L., Ferguson, K. T., & Ferguson, G. M. 2017. I am AmeriBritSouthAfrican-Zambian: Multidimensional remote acculturation and well-being among urban Zambian adolescents. *International Journal of Psychology*, 52(1): 67-76.
- Fitzsimmons, S. R., Liao, Y., & Thomas, D. C. 2017. From crossing cultures to straddling them: An empirical examination of outcomes for multicultural employees. *Journal of International Business Studies*, 48(1): 63-89.
- Friedlander, M. L., Larney, L. C., Skau, M., Hotaling, M., Cutting, M. L., & Schwam, M. 2000. Bicultural identification: Experiences of internationally adopted children and their parents. *Journal of Counseling Psychology*, 47(2): 187-198.
- Freynet, N. & Clément. 2015. Bilingualism in minority settings in Canada: Integration or assimilation? *International Journal of Intercultural Relations*, 46: 55-72.
- Fu, J. H., Chiu, C., Morris, M. W., & Young, M. J. 2007. Spontaneous inferences from cultural cues: Varying responses of cultural insiders and outsiders. *Journal of Cross-Cultural Psychology*, 38(1): 58-75.
- Gong, L. 2007. Ethnic identity and identification with the majority group: Relations with national identity and self-esteem. *International Journal of Intercultural Relations*, 31(4): 503-523.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. 1998. Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6): 1464-1480.
- Gudykunst, W. B., Matsumoto, Y., Ting-Toomey, S., Nishida, T., Kim, K., & Heyman, S. 1996. The influence of cultural individualism-collectivism, self construals, and individual values on communication styles across cultures. *Human Communication Research*, 22(4): 510-543.
- Hammack, P. L. 2010. Narrating hyphenated selves: Intergroup contact and configurations of identity among young Palestinian citizens of Israel. *International Journal of Intercultural Relations*, 34(4): 368-385.
- Hanek, K. J., Lee, F., & Brannen, M. Y. 2014. Individual differences among global/multicultural individuals. *International Studies of Management and Organization*, 44(2): 75-89.
- Hoersting, R. C., & Jenkins, S. R. 2011. No place to call home: Cultural homelessness, self-esteem and cross-cultural identities. *International Journal of Intercultural Relations*, 35(1): 17-30.
- Hsu, L. 2011. Linguistic intergroup bias tells ingroup/outgroup orientation of bicultural Asian Americans. *International Journal of Intercultural Relations*, 35(6): 853-866.
- Huff, S., Lee, F., & Hong Y-y. 2017. Bicultural and generalized identity integration predicts interpersonal tolerance. *Journal of Cross-Cultural Psychology*, 48(5): 644-666.
- Jugert, P., Rutland, A., Brown, R., Cameron, L., Nigur, D., Watters, C., Hossain, R., Landau, A., & Le Touze, D. 2017. Increasing ethnic diversity moderates longitudinal effects of individual differences on friendship homophily. *Journal of Community & Applied Psychology*, 27(5): 411-423.
- Kane, A. & Levina, N. 2017. 'Am I still one of them?': Bicultural immigrant managers navigating social identity threats when spanning global boundaries. *Journal of Management Studies*, 54(4): 540-577.
- Kanno, Y. 2000. Kikokushijo as bicultural. *International Journal of Intercultural Relations*, 24(3): 361-382.
- Kim, S. Y., & Hou, Y. 2016. Intergenerational transmission of tridimensional cultural orientations in Chinese American families: The role of bicultural socialization. *Journal of Youth and Adolescence*, 45(7): 1-14.
- Kim, S. Y., Shen, Y., Huang, X., Wang, Y., & Orozco-Lapray, D. 2014. Chinese American parents' acculturation and enculturation, bicultural management difficulty, depressive symptoms, and parenting. *Asian American Journal of Psychology*, 5(4): 298-306.

- Kuhn, M. H., & McPartland, T. S. 1954. An empirical investigation of self-attitudes. *American Sociological Review*, 19(1): 68–76.
- Kulis, S. S., Robbins, D. E., Baker, T. M., Denetsose, S., & Deschine Parkhurst, N. A. 2016. A latent class analysis of urban American Indian youth identities. *Cultural Diversity and Ethnic Minority Psychology*, 22(2): 215-228.
- Lee, J. H., & Church, T. A. 2017. Social anxiety in Asian Americans: Integrating personality and cultural factors. *Asian American Journal of Psychology*, 8(2), 103-114.
- Lin, X. & Malhotra, S. 2012. To adapt or not adapt: The moderating effect of perceived similarity in cross-cultural business partnerships. *International Journal of Intercultural Relations*, 36(1): 118-29.
- Lu, L. 2007. The individual-and social-oriented self views: Conceptual analysis and empirical assessment. *US-China Education Review*, 4, 1–24.
- Lu, L. 2008. The individual-oriented and social-oriented Chinese bicultural self: Testing the theory. *Journal of Social Psychology*, 148(3): 347-374.
- Luhtanen, R., & Crocker, J. 1992. A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin*, 18(3): 302–318.
- Luna, D., & Peracchio, L. A., 2001. Moderators of language effects in advertising to bilinguals: A psycholinguistic approach. *Journal of Consumer Research*, 28(2): 284-295.
- Luna, D., Ringberg, T., & Peracchio, L. A. 2008. One individual, two identities: Frame switching among biculturals. *Journal of Consumer Research*, 35(2): 279-293.
- Marks, A. K., Patton, F., & Coll, C. G. 2011. Being bicultural: A mixed-methods study of adolescents' implicitly and explicitly measured multiethnic identities. *Developmental Psychology*, 47(1): 270-288.
- Maslach, C. 1974. Social and personal bases of individuation. *Journal of Personality and Social Psychology*, 29(3): 411–25.
- Mendoza, R. H. 1989. An empirical scale to measure type and degree of acculturation in Mexican-American adolescents and adults. *Journal of Cross-Cultural Psychology*, 20(4), 372–385.
- Mok, A., & Morris, M. W. 2010a. An upside to bicultural identity conflict: Resisting groupthink in cultural ingroups. *Journal of Experimental Social Psychology*, 46(6): 1114–1117.
- Mok, A., & Morris, M. 2010b. Asian-Americans' creative styles in Asian American situations: Assimilative and contrastive responses as a function of bicultural identity integration. *Management and Organization Review*, 6(3), 371-390.
- Mok, A., & Morris, M. W. 2013. Bicultural self-defense in consumer contexts: Self-protection motives are the basis for contrast versus assimilation to cultural cues. *Journal of Consumer Psychology*, 23(2): 175–188.
- Ng, S. H. 2007. Biculturalism in multicultural Hong Kong. *Journal of Psychology in Chinese Societies*, 8, 121-140.
- Ng, S., & Houston, M. J. 2006. Exemplars or beliefs? The impact of self-view on the nature and relative influence of brand associations. *Journal of Consumer Research*, 32(4): 519-529.
- Ng, T. K., Ng, S. H., & Ye, S. 2016. Assimilation and contrast effects of culture priming among Hong Kong Chinese: The moderating roles of dual cultural selves. *Journal of Cross-Cultural Psychology*, 47(4): 540–557.
- Ng, T. K., Rochelle, T. L., Shardlow, S. M., & Ng, S. H. 2014. A transnational bicultural place model of cultural selves and psychological citizenship: The case of Chinese immigrants in Britain. *Journal of Environmental Psychology*, 40: 440-450.
- Ng, S. H., Yam, N., & Lai, J. C. L. 2007. The bicultural self of Chinese in Hong Kong. In J. H. Liu, C. Ward, A. B. I. Bernardo, M. Karasawa, & R. Fischer (Eds.), *Casting the individual in societal and cultural contexts: Social and societal psychology in Asia and the Pacific* (pp. 105-122). Seoul, South Korea: Kyoyook-Kwakak-Sa.
- Novin, S., Banerjee, R., & Rieffe, C. 2012. Bicultural adolescents' anger regulation: In between two cultures? *Cognition & Emotion*, 26(4): 577-586.
- O'Connor, P. (2011) Everyday hybridity and Hong Kong's Muslim youth. *Visual Anthropology*, 24(1-2): 203-225.
- Oswald, L. R. 1999. Culture swapping: Consumption and the ethnogenesis of middle-class Haitian immigrants. *Journal of Consumer Research*, 25(4): 303-318.
- Peñalozza, L., & Gilly, M. C. 1999. Marketer acculturation: The changer and the changed. *Journal of Marketing*, 63(3): 84-104.
- Perunovic, W. E., Heller, D., & Rafaeli, E. (2007). Within-person changes in the structure of emotion: The role of cultural identification and language. *Psychological Science*, 18(7): 607-613.
- Phinney, J. S. 1992. The multigroup ethnic identity measure: A new scale for use with diverse groups. *Journal of Adolescent Research*, 7(2): 156–176.
- Phinney, J. S., & Devich-Navarro, M. 1997. Variations in bicultural identification among African American and Mexican American adolescents. *Journal of Research on Adolescence*, 7(1): 3–32.

- Phinney, J. S. & Ong, A. D. 2007. Conceptualization and Measurement of Ethnic Identity: Current Status and Future Directions. *Journal of Counseling Psychology*, 54(3): 271-281.
- Ringberg, T., Luna, D., Reihlen, M., & Peracchio, L. 2010. Bicultural-bilinguals: The effect of cultural frame switching on translation equivalence. *International Journal of Cross-Cultural Management*, 10(1): 77-92.
- Rivera-Sinclair, E. A. 1997. Acculturation/biculturalism and its relationship to adjustment in Cuban-Americans. *International Journal of Intercultural Relations*, 21(3): 379-391.
- Roccas, S., Sagiv, L., Schwartz, S., Halevy, N., & Eidelson, R. 2008. Toward a unifying model of identification with groups: Integrating theoretical perspectives. *Personality and Social Psychology Review*, 12(3): 280–306.
- Ryder, A. G., Alden, L. E., & Paulhus, D. L. 2000. Is acculturation unidimensional or bidimensional? A head-to-head comparison in the prediction of personality, self-identity, and adjustment. *Journal of Personality and Social Psychology*, 79(1): 49–65.
- Saad, C. S., Damian, R. I., Benet-Martínez, V., Moons, W. G., & Robins, R. W. 2013. Multiculturalism and creativity: Effects of cultural context, bicultural identity, and ideational fluency. *Social Psychological and Personality Science*, 4(3): 369–375.
- Sandil, R., Robinson, M., Brewster, M. E., Wong, S., & Geiger, E. 2015. Negotiating multiple marginalizations: Experiences of South Asian LGBTQ individuals. *Cultural Diversity and Ethnic Minority Psychology*, 21(1): 76-88.
- Schwartz, S. J., Unger, J., Baezconde-Garbanati, L., Benet-Martínez, V., Meca, A., Zamboanga, B., Lorenzo-Blanco, E., Des Rosiers, S., Oshri, A., Sabet, R. F., Soto, D. W., Pattarroyo, M., Huang, S., Villamar, J. A., Lizzi, K. M., Szapocznik, J., Longitudinal trajectories of bicultural identity integration in recently immigrated Hispanic adolescents: Links with mental health and family functioning. *International Journal of Psychology*, 50(6): 440-450.
- Singelis, T. M. 1994. The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin*, 20(5): 580–591.
- Sirikantraporn, S. 2013. Biculturalism as a protective factor: An exploratory study on resilience and the bicultural level of acculturation among Southeast Asian American youth who have witnessed domestic violence. *Asian American Journal of Psychology*, 4(2): 109-115.
- Spiegler, O., & Leyendecker, B. 2017. Balanced cultural identities promote cognitive flexibility among immigrant children. *Frontiers in Psychology*, 8: 1-5.
- Stroink, M. L., & Lalonde, R. N. 2009. Bicultural identity conflict in second-generation Asian Canadians. *The Journal of Social Psychology*, 149(1): 44-65.
- Suinn, R. M., Rickard-Figueroa, K., Lew, S., & Vigil, P. 1987. The Suinn-Lew Asian self-identity acculturation scale: An initial report. *Educational and Psychological Measurement*, 47(2): 401–407.
- Szapocznik, J., Kurtines, W. M., & Fernandez, T. 1980. Bicultural involvement and adjustment in Hispanic-American youth. *International Journal of Intercultural Relations*, 4(3): 353–365.
- Szapocznik, J., Scopetta, M. A., Kurtines, W., & Aranalde, M. D. 1978. Theory and measurement of acculturation. *Interamerican Journal of Psychology*, 12(2): 113–130.
- Tadmor, C. T., Galinsky, A. D., & Maddux, W. M. 2012. Getting the most out of living abroad: Biculturalism and integrative complexity as key drivers of creative and professional success. *Journal of Personality and Social Psychology*, 103(3): 520-542.
- Tadmor, C. T., Tetlock, P. E., & Peng, K. 2009. Acculturation strategies and integrative complexity: The cognitive implications of biculturalism. *Journal of Cross-Cultural Psychology*, 40(1): 105–139.
- Tawa, J., & Suyemoto, K. L. 2010. The influence of race and power on self-construal in bicultural Asian Americans. *Asian American Journal of Psychology*, 1(4): 275-289.
- Thompson, C. J., & Tambyah, S. K. 1999. Trying to be cosmopolitan. *Journal of Consumer Research*, 26(3): 214-241.
- Ting-Toomey, S., Yee-Jung, K. K., Shapiro, R. B., Garcia, W., Wright, T. J., Oetzel, J. G. 2000. Ethnic/cultural identity salience and conflict styles in four US ethnic groups. *International Journal of Intercultural Relations*, 24(1): 47-81.
- Tropp, L. R., & Wright, S. C. 2001. Ingroup identification as the inclusion of ingroup in the self. *Personality and Social Psychology Bulletin*, 27(5): 585–600.
- Umaña-Taylor, A. J., Yazedjian, A., & Bámaca-Gómez, M. 2004. Developing the ethnic identity scale using Eriksonian and social identity perspectives. *Identity: An International Journal of Theory and Research*, 4(1): 9–38.
- van de Vijver, F. J. R., Blommaert, J., Gkoumasi, G. & Stogianni, M. 2015. On the need to broaden the concept of ethnic identity. *International Journal of Intercultural Relations*, 46: 36-46.
- Ward, C. 2006. Acculturation, identity and adaptation in dual heritage adolescents. *International Journal of Intercultural Relations*, 30(2): 243–259.

- Ward, C., & Kennedy, A. 1994. Acculturation strategies, psychological adjustment, and sociocultural competence during cross-cultural transitions. *International Journal of Intercultural Relations*, 18(3): 329–343.
- Ward, C., Stuart, J., & Kus, L. 2011. The construction and validation of a measure of ethno-cultural identity conflict. *Journal of Personality Assessment*, 93(5), 462-473.
- Wei, M., Liao, K. Y., Chao, R. C., Mallinckrodt, B., Tsai, P., & Botello-Zamarron, R. 2010. Minority stress, perceived bicultural competence, and depressive symptoms among ethnic minority college students. *Journal of Counseling Psychology*, 57(4): 411-422.
- Wiley, S. 2013. Rejection-identification among Latino immigrants in the United States. *International Journal of Intercultural Relations*, 37(3): 375-384.
- Yamada, A.-M., & Singelis, T. M. 1999. Biculturalism and self-construal. *International Journal of Intercultural Relations*, 23(5): 697–709.
- Yampolsky, M. A., Amiot, C. E., & de la Sablonnière, R. 2013. Multicultural identity integration and well-being: A qualitative exploration of variations in narrative coherence and multicultural identification. *Frontiers in Psychology*, 4: 1–15.
- Yampolsky, M. A., Amiot, C. E., & de la Sablonnière, R. 2016. The multicultural identity integration scale (MULTIIS): Developing a comprehensive measure for configuring one’s multiple cultural identities within the self. *Cultural Diversity and Ethnic Minority Psychology*, 22(2): 166-184.
- Zak, I. 1973. Dimensions of Jewish-American identity. *Psychological Reports*, 33(3): 891–900.
- Zea, M. C., Asner-Self, K. K., Birman, D., & Buki, L. P. 2003. The abbreviated multidimensional acculturation scale: Empirical validation with two Latino/Latina samples. *Cultural Diversity and Ethnic Minority Psychology*, 9(2): 107–126.
- Zhang, Y., & Khare, A. 2009. The impact of accessible identities on the evaluation of global versus local products. *Journal of Consumer Research*, 36(3): 524-537.