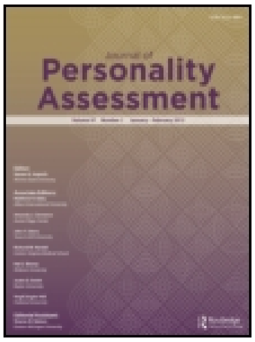

Journal of Personality Assessment

ISSN: 0022-3891 (Print) 1532-7752 (Online) Journal homepage: http://www.tandfonline.com/loi/hjpa20

Examining the Process of Responding to Circumplex Scales of Interpersonal Values Items: Should Ideal Point Scoring Methods Be Considered?

Ying Ling, Minqiang Zhang, Kenneth D. Locke, Guangming Li & Zonglong Li

JOURNAL OF PERSONALITY ASSESSMENT
2016, VOL. 98, NO. 3, 310–318
http://dx.doi.org/10.1080/00223891.2015.1077852

Routledge
Taylor & Francis Group

Examining the Process of Responding to Circumplex Scales of Interpersonal Values Items: Should Ideal Point Scoring Methods Be Considered?

Ying Ling,[1] Minqiang Zhang,[1] Kenneth D. Locke,[2] Guangming Li,[1] and Zonglong Li[1]

[1]Center for Studies of Psychological Application, School of Psychology, South China Normal University, Guangzhou, China; [2]Department of Psychology, University of Idaho

ABSTRACT
The Circumplex Scales of Interpersonal Values (CSIV) is a 64-item self-report measure of goals from each octant of the interpersonal circumplex. We used item response theory methods to compare whether dominance models or ideal point models best described how people respond to CSIV items. Specifically, we fit a polytomous dominance model called the generalized partial credit model and an ideal point model of similar complexity called the generalized graded unfolding model to the responses of 1,893 college students. The results of both graphical comparisons of item characteristic curves and statistical comparisons of model fit suggested that an ideal point model best describes the process of responding to CSIV items. The different models produced different rank orderings of high-scoring respondents, but overall the models did not differ in their prediction of criterion variables (agentic and communal interpersonal traits and implicit motives).

ARTICLE HISTORY
Received 7 November 2014
Revised 29 May 2015

In recent decades, the interpersonal circle or interpersonal circumplex (IPC) has become the most popular model for conceptualizing, organizing, and assessing interpersonal dispositions (Wiggins, 2003). The IPC is defined graphically by two orthogonal axes: a vertical axis (of status, dominance, power, control, or, most broadly, agency) and a horizontal axis (of solidarity, friendliness, warmth, love, or, most broadly, communion). The circumplex is typically divided into the eight octants, each labeled with a generic two-letter code (shown in parentheses in Figure 1). As one circumnavigates the circle, each octant reflects a blend of agency and communion; for example, the NO octant reflects high agency and high communion, whereas the BC octant reflects high agency and low communion.

IPC inventories are inventories designed to measure interpersonal dispositions from every segment of the IPC (Locke, 2011). There exist IPC measures of many different interpersonal constructs, including interpersonal traits, interpersonal problems, interpersonal self-efficacy, and interpersonal sensitivities. One such inventory is the Circumplex Scales of Interpersonal Values (CSIV; Locke, 2000).

Circumplex scales of interpersonal values

The CSIV consists of eight 8-item scales that assess the value individuals place on agentic and communal interpersonal outcomes or modes of conduct associated with each IPC octant (e.g., "I feel connected to them," "I keep my guard up"). For each item, respondents indicate how important that type of interpersonal experience is for them on a 5-point scale (*not important, mildly important, moderately important, very important, extremely important*). The CSIV scales have demonstrated acceptable levels of internal consistency, test–retest reliability, circumplex structure, and convergent and discriminant validity with measures of interpersonal traits, interpersonal problems, interpersonal goals, and interpersonal sensitivities (Hopwood et al., 2011; Locke, 2000). Because the preceding studies employed self-report measures, it is worth noting that the CSIV scales also correlate with implicit or indirect measures of interpersonal motives (Locke, 2000; Turan & Horowitz, 2010) and even testosterone levels in blood samples (Turan, Guo, Boggiano, & Bedgood, 2014).

The CSIV has been used in numerous research studies. For example, stronger communal values (as assessed by the CSIV) have been shown to predict (a) perceiving others as more similar to the self (Locke & Christensen, 2007; Locke, Craig, Baik, & Gohil, 2012), (b) experiencing more positive emotions when noticing similarities between oneself and others (Locke, 2003), (c) harsher judgments of perpetrators of antisocial actions (Kammrath & Scholer, 2011), and (d) greater dyadic satisfaction when two people work together on a task (Locke & Sadler, 2007). Regarding clinical applications, research has shown that (a) different personality disorder symptoms predict distinct patterns of CSIV scores (Locke, 2000), (b) individuals with a clear pattern of CSIV scores—with high scores in one region of the IPC and low scores in the opposite region—tend to make interpersonal decisions more easily (Locke & Adamic, 2012), and (c) in psychosomatic patients, unagentic values are

CONTACT Minqiang Zhang zhangmq1117@qq.com Center for Studies of Psychological Application, School of Psychology, South China Normal University, Zhongshan West Road, Guangzhou, 510631, China.
Supplemental material for this article can be accessed on the publisher's website.

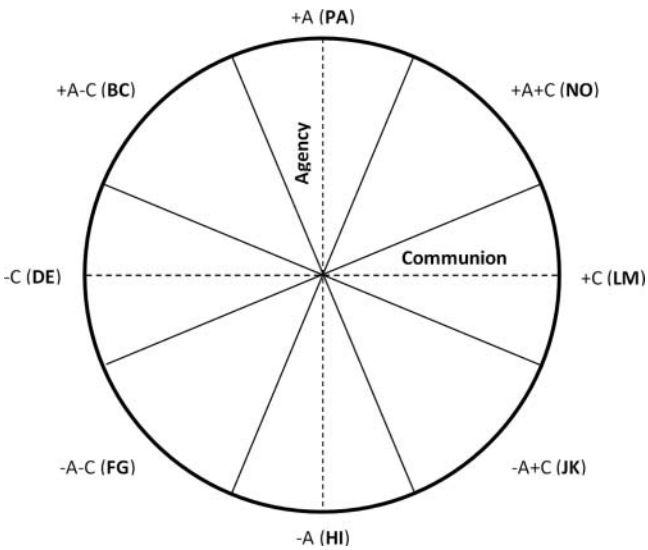

Figure 1. The eight octants of the interpersonal circumplex.

associated with greater distress and also tend to decline over the course of treatment (Thomas, Kirchmann, Suess, Brautigam, & Strauss, 2012). Thus, the CSIV can be a useful tool in research, clinical, and educational settings; however, maximizing the utility of the CSIV requires maximizing the accuracy with which the octant scores reflect the underlying dispositions they purport to measure.

Dominance versus ideal point item response processes

The assumptions that scale developers make about how people respond to items influence how they combine those responses into scale scores. Consequently, misspecification of the response process can adversely affect the accuracy of trait scores and the predictions made concerning respondents' behavior (Stark, Chernyshenko, Drasgow, & Williams, 2006).

The assumption of a dominance response process is that a respondent will tend to endorse an item to the degree that the respondent's location on the latent dimension exceeds the location of the item on that dimension, and thus the probability of observing a high item score increases monotonically with increasing positive distance between the location of the respondent and the location of the item (Stark et al., 2006). The function or curve relating the latent trait level—denoted by the Greek letter theta (θ)—and the probability of item endorsement is called the item response function or item characteristic curve (ICC). Figure 2 shows an ICC for a hypothetical item that fits a dominance response process (specifically, an ICC for the generalized partial credit model [GPCM]). For example, if this was the ICC for a CSIV item that describes a moderately submissive interpersonal experience, then a person who values being extremely submissive would have a higher probability of a positive response than would either a person who values being moderately submissive or a person who values being dominant.

The assumption of an ideal point response process is that a respondent will tend to endorse an item to the degree that the respondent's perception of his or her location on the latent dimension (i.e., the respondent's *ideal point*) is similar to the location of the item on that dimension, and thus the probability of observing a high item score decreases with increasing distance between the location of the respondent and the location of the item (Stark et al., 2006). Because the respondent might avoid endorsing an item because he or she is located either too far above or too far below the item, the resulting ICC will be a nonmonotonic function with a single peak. Figure 3 shows an ICC for a hypothetical item that fits the assumptions of an ideal point response process (specifically, an ICC for the generalized graded unfolding model [GGUM]). For example, if this was the ICC for a CSIV item that describes a moderately submissive interpersonal experience, then a person who values being moderately submissive would have a higher probability of a positive response than would either a person who values being extremely submissive or a person who values being dominant.

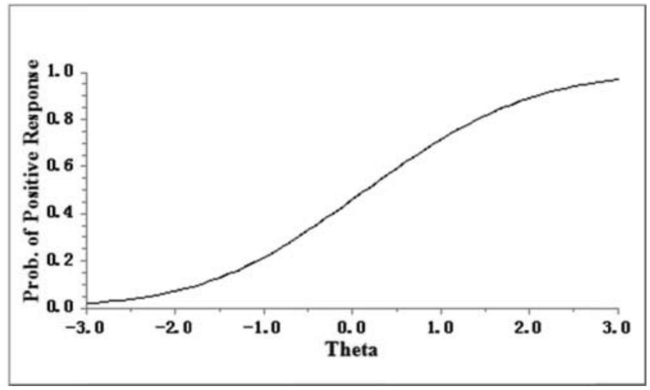

Figure 2. Example of item characteristic curve for a dominance response process model (generalized partial credit model). The probability of positive response is the likelihood of endorsing the most positive response option.

Advantages of IRT and ideal point models

Item response theory (IRT) refers to model-based measurement in which trait-level estimates depend on both the items' properties and the individuals' responses (Embretson & Reise, 2000). The estimate of the person parameter—the "score" on a test when employing IRT—can offer advantages over traditional scores based on classical test theory (CTT). Some benefits of IRT are most relevant to ability and achievement testing, but

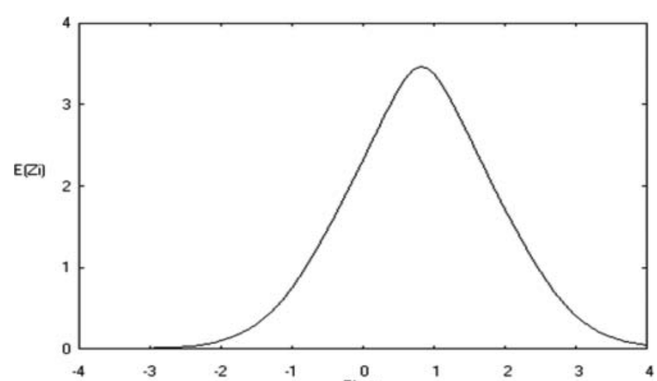

Figure 3. Example of item characteristic curve for an ideal point response process model (generalized graded unfolding model). E(Zi) is the expected item response averaged over response categories (here assumed to range from 0–4).

several of the advantages are relevant to personality tests like the CSIV. One such advantage is that changes in the particular items sampled (e.g., sampling more strongly endorsed items) or persons sampled (e.g., sampling more individuals with strong communal values) might influence CTT estimates of person and item parameters, but not the corresponding IRT estimates; therefore, using IRT facilitates the comparison of latent traits and item properties across different populations and situations (e.g., across cultures). Another advantage is that IRT yields an optimal scaling of individual differences, and whereas CTT assumes measurement error is a property of the scale (equivalent for all respondents), IRT allows measurement error to vary across the latent trait continuum; for example, a scale might show reduced measurement precision (discriminating power) among respondents at the upper extreme of the trait distribution. Yet another advantage is that whereas CTT estimates person and item parameters on different scales, IRT estimates them on the same (latent trait) scale. Perhaps most relevant to this article is that research does suggest that IRT methods can improve the accuracy and precision of circumplex scales (Sodano & Tracey, 2011; Wetzel & Hell, 2014).

However, the sundry advantages of IRT, which distinguish it from CTT, depend on accurate specification of the item response process (Stark et al., 2006). Intuitively, many of the items on personality scales like the CSIV (e.g., how important is it that others not get their feelings hurt) should fit an ideal point better than a dominance response process. Moreover, there is evidence that ideal point models do provide a more accurate functional specification of the trait−response relationship and more accurate rank orderings of individuals' latent traits for at least some noncognitive measures (Carter & Dalal, 2010; Chernyshenko, Stark, Drasgow, & Roberts, 2007; Roberts, Donoghue, & Laughlin, 2000; Stark et al., 2006). In this article we examine empirically whether responses to CSIV items are best described by an ideal point or a dominance response process.

Testing item response processes of measures with IRT

As described next, IRT offers both graphical and statistical methods of testing the item response process. The graphical method involves examining ICCs. The statistical approach involves comparing the model fit for dominance and ideal point models.

Graphically examining ICCs

Folding in ICCs provides visible support for the ideal point response process assumed by ideal point models. Specifically, whereas dominance models expect a monotonically increasing ICC, ideal point models expect a nonmonotonic ICC that peaks around the point of the item's estimated location, *delta* (δ_i), on the trait continuum. However, the characteristic peaking and folding will only be evident for items with an estimated location that is within the typical range of thetas. For example, if ICCs are plotted over the thetas ranging from −4 to +4, then folding would not be observed for items with more extreme location parameters of $|\delta_i| > 4$. Theoretically, extreme items might show folding if ICCs were plotted over a broader range of trait levels (e.g., 6 *SD*); however, empirically this is unlikely because few if any individuals have trait levels (ideal points) that deviate more than 4 *SD* from the mean (Stark et al., 2006). Thus, for extreme items, because almost all respondents' trait levels are not located within the range in which folding occurs, ideal point and dominance models will fit response data similarly well, despite the difference in their basic assumptions (Stark et al., 2006). To address this issue, in our analyses we categorized items as very extreme ($|\delta_i| > 4$), moderately extreme ($4 > |\delta_i| > 2$), or relatively neutral ($|\delta_i| < 2$). We expected the ideal point ICCs of very extreme items to show no folding because the location of the item on the trait dimension exceeds that of almost all respondents; the ICCs of moderately extreme items to show some but not strong folding because the location of the item exceeds that of most respondents; and the ICCs of relatively neutral items to show strong folding because the location of the item on the trait dimension is within the same range as the location of most respondents [−2,+2].

Statistically comparing model fit

Another way to evaluate response processes involves statistically comparing the model−data fit for the dominance models and the ideal point models. A dominance model could be considered a special case of an ideal point model, because whereas dominance IRT models assume monotonically increasing ICCs, ideal point models can model both monotonic and nonmonotonic ICCs (Stark et al., 2006). Therefore, to the degree the data are nonmonotonic, ideal point models should fit the data better than the dominance models. Specifically, we can compute χ^2/df ratios for singles, pairs, and triples of items within one subscale for both the dominance models and the ideal point models (for details of the formulas used to compute the χ^2/df ratios, see the online supplemental materials). Item singles are a measure of the difference between the observed scores in the data and the scores that would be expected by the IRT model. Doubles and triples are sensitive to violations of unidimensionality and local independence (Glas, 1988; Van den Wollenberg, 1982). To assess relative model−data fit, we can also calculate the Akaike's information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 1978); both AIC and BIC permit comparisons of model likelihoods while penalizing more complex models.

This study

Like other IPC measures, the CSIV was developed based on CTT methods and dominance response process assumptions. However, responding on the CSIV—which asks respondents to select which option best describes their values—might be more consistent with an ideal point response process than with a dominance response process. Indeed, several investigations suggest that ideal point response process assumptions can be an effective alternative for scale development and scoring of measures of noncognitive attributes such as attitudes and personality traits (Roberts, Laughlin, & Wedell, 1999; Stark et al., 2006). However, there is no research examining which model of item response process is most applicable to either IPC measures or measures of goals, such as the CSIV. Therefore, the purpose of this study was to employ IRT methods to compare whether a dominance model or an ideal point model is most accurate in describing and most useful in scoring responses to the CSIV. Specifically, we compared a polytomous dominance IRT model called the GPCM (Muraki, 1992) with an ideal point IRT model of similar complexity called the GGUM (Roberts et al., 2000).

We used both graphical and statistical methods to assess the item response processes of CSIV with IRT. The graphical approach involves examining the shape of the GGUM ICCs. The hypothesis (H1) was that, to the extent that CSIV items fit an ideal point response process, the GGUM ICCs would be nonmonotonic, and single-peaked for items that contained relatively neutral content. The statistical approach involves comparing the model fit for the two models—the GGUM and the GPCM. The hypothesis (H2) was that, to the extent that CSIV items fit an ideal point response process, the GGUM would show better model−data fit than the GPCM.

Even if an ideal point model better describes CSIV responses, the practical question remains: What difference does it make? The literature on structural validity clearly illustrates the issue: On the one hand, personality items rarely map onto a personality inventory's theoretical structure with sufficient fidelity to meet conventional confirmatory factor analysis criteria; on the other hand, these deviations from the ideal might not undermine criterion-related validity relative to using a more complex model that mirrors the interrelationships among the items more precisely (Herrmann & Pfister, 2013; Hopwood & Donnellan, 2010). However, the impact of misspecifications of scale structure (which is tested using factor analytic methods under dominance process assumptions) might be distinct from the impact of misspecifications of item response processes. Therefore, we also tested the effect of measurement models on the placement of respondents and on convergent validity. We hypothesized that traditional (CTT) and ideal point (GGUM) models would diverge more in their rank ordering of respondents at the upper end of the latent distributions (H3), and that GGUM thetas would predict validity criteria more strongly than traditional CTT scores (H4). Whereas the first three hypotheses concern the functioning of items within each octant (treated as separate, unidimensional scales), the fourth hypothesis concerns functioning of the inventory (with the correlated scales organized into a circumplex).

Method and results

Participants and data analysis

The participants were undergraduates ($N = 1{,}894$) enrolled in various psychology or communications courses at the University of Idaho. Specifically, the data were CSIV responses from previously published studies (Locke, 2000, 2003; Locke & Christensen, 2007; Locke et al., 2012) and one unpublished study. To assess convergent validity, we used Locke's (2000) Bem Sex Role Inventory (BSRI; Bem, 1974) and Thematic Apperception Test (TAT) data; the BSRI was used as a measure of agentic and communal traits; the TAT was used as a measure of implicit needs for power and intimacy (see Locke, 2000, for details). We analyzed the data using the programs SPSS 13.0, GGUM2004 (Roberts, Fang, Cui, & Wang, 2006; Roberts & Shim, 2008), PARSCALE 4.1 (Muraki & Bock, 2003), and MODFIT (Stark, 2001), and computed effect sizes using online calculators available at www.psychometrica.de.

Testing item response processes of CSIV

Assessing unidimensionality of CSIV octants

Assessing unidimensionality is necessary to determine the appropriateness of the GPCM. Therefore, we conducted component analyses to assess the unidimensionality of each CSIV octant scale. Using Reckase's (1979) recommendation that the first factor of a measure should account for at least 20% of the measure's total variance, all eight scales had sufficient unidimensionality to meet the assumption of the GPCM; specifically, the total variance explained by the first component was 36%, 43%, 43%, 44%, 42%, 43%, 49%, and 39%, respectively, for the PA, BC, DE, FG, HI, JK, LM, and NO scales. Therefore, it was appropriate to use the GPCM for these data. However, these methods are inappropriate for judging whether the data are appropriate for the GGUM. We judged that by the model fit of GGUM for CSIV.

CSIV GGUM ICCs

Three items (#41 [PA scale], #16 [JK], and #06 [NO scale]) had extreme locations ($|\delta_i| > 4$) and showed no folding; 18 items (#28 [BC], #15, #55, #63 [DE], #10, #26, #34, #42, #50, #58 [FG], and all 8 HI items) were relatively neutral ($|\delta_i| < 2$) and showed strong folding; and the remaining 43 items were moderately extreme ($4 > |\delta_i| > 2$) and showed some folding (see online supplemental Table A.1). Because the items within each category showed similar GGUM ICCs, Figure 4 shows ICCs for only one item from each category simply to illustrate the typical pattern observed for the items in that category. (The GGUM ICCs for all of the CSIV items are shown in online supplemental Figures A.1 through A.8). To summarize the graphical findings, 3 CSIV items showed very extreme locations and GGUM ICCs that lacked folding, but the other 61 CSIV items showed less extreme locations with relatively neutral item content and GGUM ICCs that were nonmonotonic and single-peaked, reflecting an ideal point response process.

Statistical comparisons of model fit between GGUM and GPCM

Table 1 presents a summary of the model−data fit results for each CSIV scale. The first three columns show the number of ICCs that exhibited folding according to the GGUM, and confirm that almost all items in every scale show some folding. Columns 4 and 5 show adjusted χ^2/df ratios. Values for item singles, doubles, and triples were averaged over items to obtain an overall fit index for each scale. (Online supplemental Table A.3 reports each scale's adjusted χ^2/df ratios separately for item singles, doubles, and triples). The adjusted χ^2/df ratios averaged across items of GGUM (2.35) were less than that of GPCM (6.37), indicating better data fit for the GGUM. Columns 6 through 9 show the AIC and BIC for the GGUM and GPCM models; consistent with the adjusted χ^2/df ratios, AIC and BIC were smaller for GGUM, again indicating that GGUM was the better fitting model.

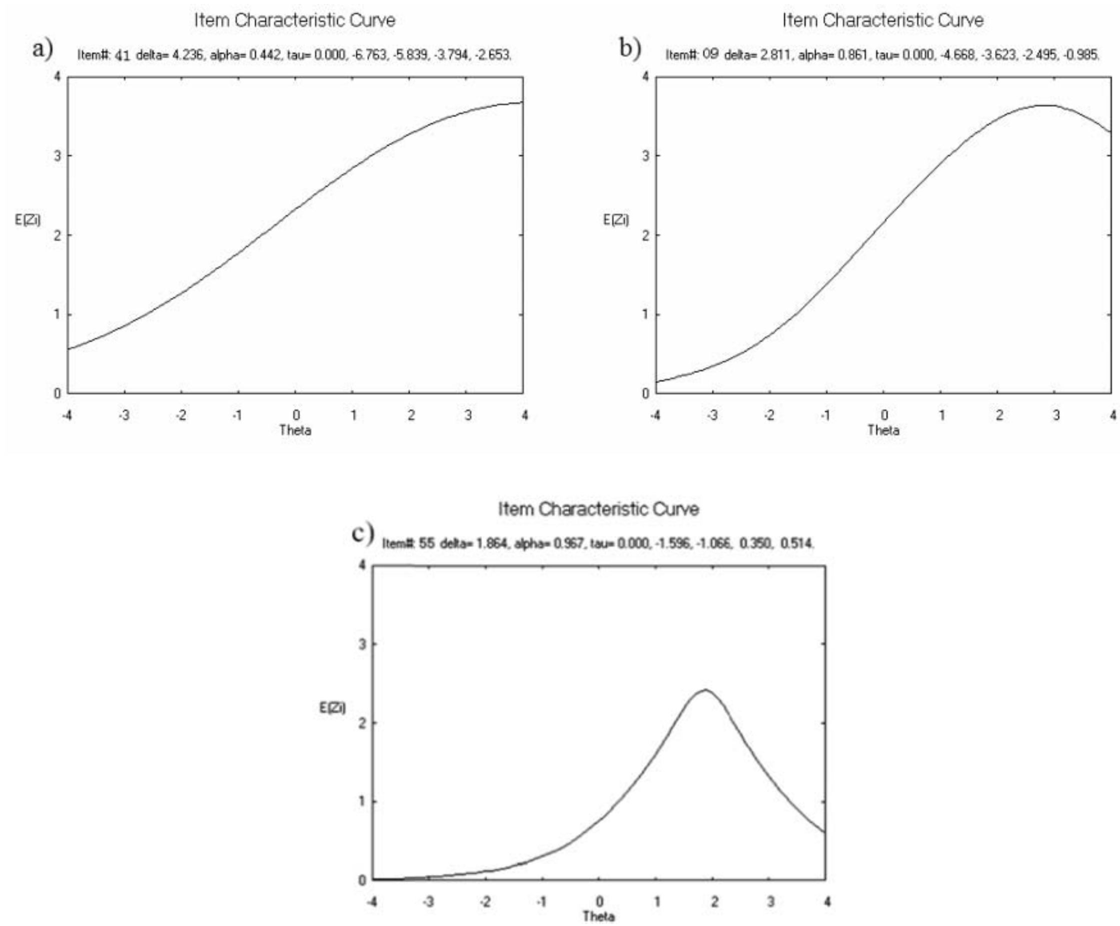

Figure 4. Examples of generalized grade unfolding model item characteristic curves of Circumplex Scales of Interpersonal Values items showing (a) no folding, (b) some folding, and (c) clear folding. E(Zi) is the expected item response averaged over response categories (which ranged from 0−4).

In summary, although the CSIV was constructed by methods assuming a dominance process, the GGUM ICCs show the single-peaked, nonmonotonic curves predicted by an ideal point response process, in accord with our hypothesis (H1). Statistical comparisons confirmed that GGUM showed better model−data fit than the GPCM, supporting our hypothesis (H2) that responding to CSIV items reflects an ideal point process.

Effect of measurement models on rank-ordering of respondents

We examined the relationships between CTT "true scores" and GGUM trait levels by means of GGUM test characteristic curves. A test characteristic curve portrays the expected total score (a CTT "true score") for each value of the GGUM trait parameter (θ). Figure 5 shows test characteristic curves for each

Table 1. Numbers of items showing folding (using GGUM) and indexes of model fit (using GPCM or GGUM).

	No. items showing no, some, and strong folding			Indexes of model fit					
				Adjusted χ^2/df ratios		AIC		BIC	
CSIV octant scale	No	Some	Strong	GPCM	GGUM	GPCM	GGUM	GPCM	GGUM
Agentic (PA)	1	7	0	4.03	1.71	41124.30	16323.42	41346.16	16951.88
Agentic and separate (BC)	0	7	1	5.78	2.11	39798.98	15053.80	40020.84	15680.52
Separate (DE)	0	5	3	7.19	1.97	36744.08	12214.13	36965.94	12839.18
Submissive and separate (FG)	0	2	6	9.21	3.66	40833.59	16132.15	41055.45	16759.54
Submissive (HI)	0	0	8	4.37	3.14	41922.79	17131.92	42144.64	17759.93
Submissive and communal (JK)	1	7	0	6.25	2.36	38731.28	13830.86	38953.14	14459.16
Communal (LM)	0	8	0	9.50	1.87	35584.31	10359.17	35806.17	10987.63
Agentic and communal (NO)	1	7	0	4.62	1.96	35492.81	10571.35	35714.67	11199.81
Average				6.37	2.35	38779.02	13952.10	39000.88	14579.71

Note. GGUM = generalized graded unfolding model; GPCM = generalized partial credit model; CSIV = Circumplex Scales of Interpersonal Values; AIC = Akaike's information criterion; BIC = Bayesian information criterion. Adjusted χ^2/df ratios: values for item singles, doubles, and triples were averaged over items to obtain an overall fit index for each scale.

CSIV octant. The curves show that in the JK, LM, NO, PA, and BC octants, true scores increase as θ increases across almost the entire range of theta. However, in the DE, FG, and HI octants, true scores only increase as θ increases up to a certain level; among high scorers, as θ increases, true scores decrease.

To clarify the source of the problem, we rank-ordered respondents based on their GGUM thetas, and then computed product−moment correlations between GGUM and CTT scores for the top 100, 200, 300, and 500 scorers (as well as for the total sample of 1,894 respondents). As Table 2 shows, using

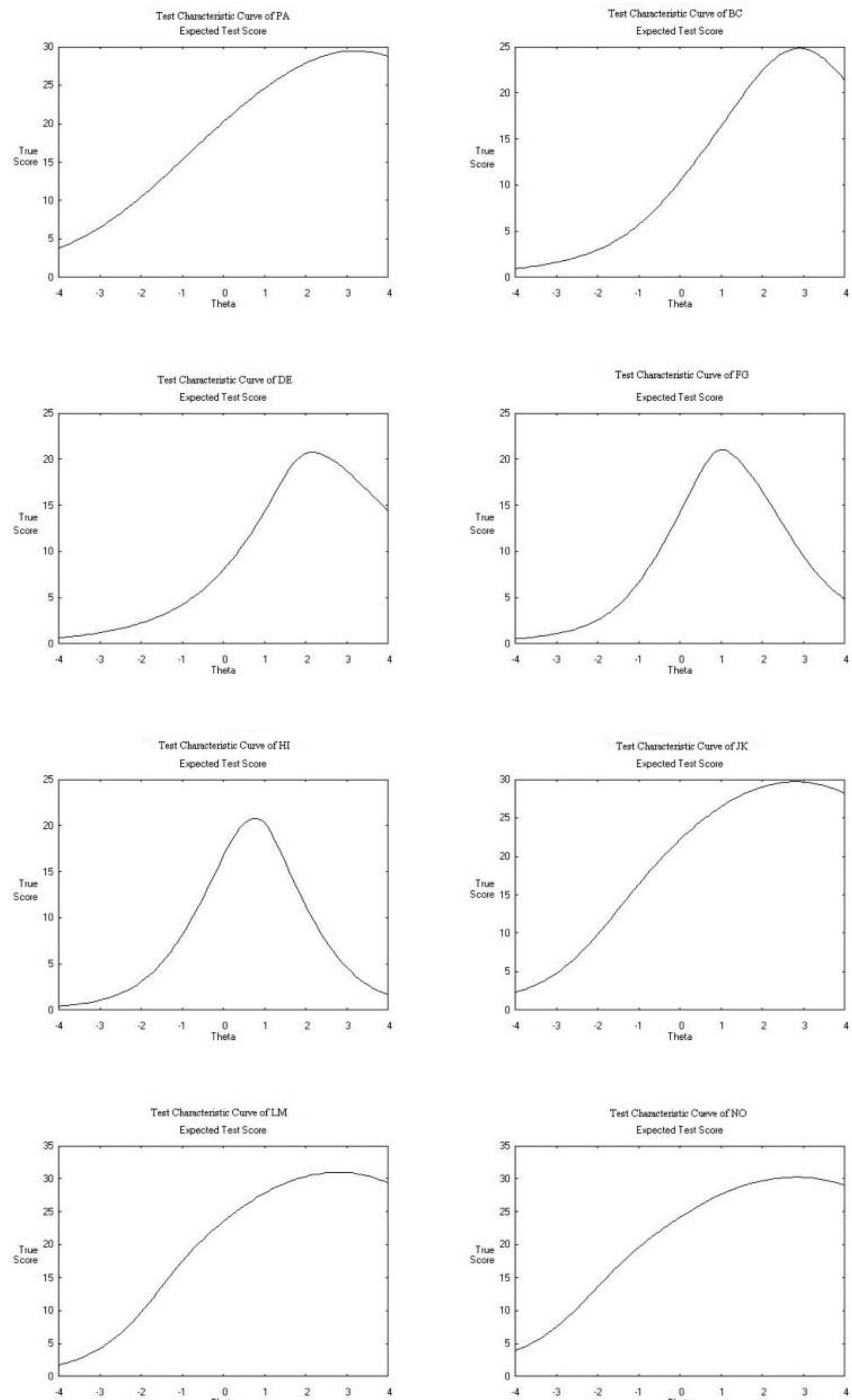

Figure 5. Generalized graded unfolding model test characteristic curves for each octant of the Circumplex Scales of Interpersonal Values.

the complete sample yielded strong correlations, but in the subsamples of high-scoring respondents the correlations were weaker, with the effect being particularly dramatic in the DE, FG, and HI (low agency and low communion) octants. Indeed, in the FG and HI octants there was no agreement in the rank-ordering of respondents using GGUM versus CTT. Therefore, in accord with our hypothesis (H3), misspecification of the CSIV response process can alter the rank order of high-scoring individuals, especially in those octants characterized by items that do not invite extreme responses.

Effect of measurement models on measures of validity

Next, we tested if using the GGUM measurement model influenced the ability of the CSIV to predict other variables—specifically, BSRI dominance and nurturance (i.e., agentic and communal traits) and TAT power and intimacy (i.e., implicit agentic and communal motives).

Our preceding analyses concerned how people respond to the items within each octant scale, and treated each octant as a separate, unidimensional scale. However, a distinctive appeal of circumplex inventories is that the interrelationships among the octant scales conform to a specific, circumplex pattern (Gurtman & Pincus, 2003). To formally test if the CSIV octant scores fit a circumplex pattern, we used the package CircE (Grassi, Luccio, & Di Blas, 2010) in R (R Core Team, 2013), with the octants constrained to have equal communality indexes and equal spacing around the circle; we only used cases ($n = 265$) that had GGUM scores for all octants. Using conventional criteria (Gurtman & Pincus, 2003), both GGUM octant scores and CTT octant scores showed good fit to a circumplex structure: for GGUM, root mean square error of approximation (RMSEA) = .084, adjusted goodness-of-fit index (AGFI) = .939; for CTT, RMSEA = .065, AGFI = .963.

When octant scales form a circle, trigonometric formulas can be used to combine the eight octants into an overall horizontal or communal dimension score (reflecting a respondent's overall communal vs. uncommunal tendencies) and an overall vertical or agentic dimension score (reflecting a respondent's overall agentic vs. unagentic tendencies) as follows (Wiggins, 2003):

$$\text{Communal Dimension} = \text{LM} - \text{DE} + (.707 * (\text{JK} + \text{NO} - \text{BC} - \text{FG})), \quad (1)$$

$$\text{Agentic Dimension} = \text{PA} - \text{HI} + (.707 * (\text{BC} + \text{NO} - \text{JK} - \text{FG})). \quad (2)$$

When predicting outcomes, it is often preferable to use dimension scores rather than octant scores because dimensions scores are more parsimonious (reducing the number of predictors from eight to two), more stable, and unaffected by overall response elevation. Therefore, we correlated the BSRI and TAT scores with the GGUM and CTT communal and agentic dimension scores.

Table 3 shows the results. As in Locke (2000), BSRI nurturance traits and TAT intimacy needs correlated positively with the CSIV communion dimension; BSRI dominance traits correlated positively with the CSIV agency dimension; and TAT power needs correlated positively with the CSIV agency dimension and negatively with the CSIV communion dimension. However, there were no significant differences—using Cohen's (1988) q test—between the correlations yielded by the CTT model versus the GGUM model.

We also correlated the BSRI and TAT scores with the GGUM and CTT for each octant scale. The results paralleled those in Table 3: There were no significant differences between the criterion-related validity coefficients for GGUM thetas versus CTT scores (online supplemental Table A.4 reports the results). Therefore, the results did not support our hypothesis (H4) that criterion variables would relate more strongly to GGUM thetas than to CTT scores.

Table 2. Correlations between GGUM latent trait scores and CSIV octant scores for different subgroups of participants.

	Correlation							
Participant subgroup	PA	BC	DE	FG	HI	JK	LM	NO
Top 100	.84	.81	.56	−.17	.03	.74	.81	.82
Top 200	.82	.82	.68	−.08	−.10	.76	.76	.76
Top 300	.85	.84	.75	−.01	−.11	.76	.77	.78
Top 500	.88	.86	.83	.17	−.10	.80	.81	.75
All 1,894	.97	.95	.96	.84	.71	.95	.96	.89

Note. GGUM = generalized grade unfolding model; CSIV = Circumplex Scales of Interpersonal Values; PA = agentic; BC = agentic and separate; DE = separate; FG = submissive and separate; HI = submissive; JK = submissive and communal; LM = communal; NO = agentic and communal.

Discussion

The CSIV was developed using CTT and assuming a dominance response process; that is, respondents were presumed to endorse an item if their location on the trait continuum exceeded that of the item. However, multiple studies suggest that ideal point response processes might better describe how individuals approach personality assessments (Stark et al., 2006). This study

Table 3. Effect of measurement models on criterion-related validity of CSIV agentic and communal dimension scores.

	BSRI		TAT	
Dimension	Dominance	Nurturance	Power	Intimacy
Communion				
GGUM	.09	.53**	−.13	.18*
CTT	.08	.51**	−.14	.18*
q		.05		.00
Agency				
GGUM	.31**	−.11	.16*	−.02
CTT	.27**	−.11	.17**	−.03
q	.04		.01	

Note. CSIV = Circumplex Scales of Interpersonal Values; BSRI = Bem Sex Role Inventory; TAT = Thematic Apperception Test; GGUM = generalized grade unfolding model; CTT = classical test theory; Communal dimension = LM − DE + (.707 * (JK + NO − BC − FG)); Agentic dimension = PA − HI + (.707 * (BC + NO − JK − FG)). q = Cohen's (1988) q test, an effect size measure that interpreted the differences between significant correlations with GGUM theta estimates and significant correlations with CTT scores, $q < .1$ = no effect. Due to invariant patterns identified by the program GGUM2004, we omitted 10 participants from the Communal-BSRI analyses (final $n = 74$), 6 participants from the Agentic-BSRI analyses (final $n = 78$), and 8 participants from the TAT analyses (final $n = 192$).

* $p < .05$.

** $p < .01$.

investigated whether the assumptions of an ideal point response process or those of a traditional dominance response process were more applicable to CSIV items. Specifically, we used IRT methods to graphically and statistically evaluate the item response process, and whether ideal point and classical scoring procedures yield different rank orderings of respondents and different correlations with criterion variables.

The results indicate that the GGUM fit the CSIV data well. First, in accord with H1, for 61 of the 64 items (i.e., those expressing relatively neutral content), the GGUM produced nonmonotonic, single-peaked curves reflecting an ideal point response process. The three items that had monotonic GGUM ICCs also had very extreme item location parameters ($|\delta_i| > 4$); items with extreme locations are a priori expected tend to yield monotonic ICCs and to be described similarly by ideal point and dominance models (Stark et al., 2006), and that expectation was clearly met in the current data. Second, in accord with H2, the GGUM (ideal point response process model) showed better fit to CSIV data than did the GPCM (dominance response process model). In sum, although the three very extreme items were fitted similarly well both by ideal point models and dominance models, the remaining 61 items were fitted better by GGUM than GPCM, suggesting an ideal point response process.

In general, people tend to express stronger agentic (+A) than unagentic (−A) values, and stronger communal (+C) than uncommunal (−C) values. Therefore, respondents were much more likely to endorse options at the upper (very important or extremely important) end of the response scale when responding to CSIV items from the PA, NO, LM, and JK octants (reflecting agentic or communal values) than when responding to CSIV items from the BC, DE, FG, and HI octants (reflecting unagentic and uncommunal values). As a consequence, the GGUM item location parameters from the PA, NO, LM, and JK octants were all at least moderately extreme ($|\delta_i| > 2$). The tendency for respondents to endorse items with agentic and communal content more than items with unagentic and uncommunal content occurs on other IPC inventories as well, potentially yielding extreme locations at the high end for agentic and communal items or at the low end for unagentic and uncommunal items. However, to the extent that most items on most IPC inventories do not express very extreme content, the item locations will also not be very extreme, and an ideal point model might be more appropriate for those inventories as well.

In accord with our third hypothesis (H3) and previous findings (Chernyshenko et al., 2007; Stark et al., 2006), we found noteworthy divergences in the rank-ordering of GGUM latent trait scores and average scale scores among high-scoring individuals. Divergences were found in all octants, because all octants contained items that showed folding; however, the divergences were particularly striking—including some negative correlations—in the FG and HI octants, which were the octants containing the most items showing strong folding.

Finally, the results provided no support for the hypothesis (H4) that criterion-related validity would be greater for GGUM thetas than traditional CTT mean scores. The ability of the CSIV octant scales to predict agentic and communal traits or implicit motives did not depend on which scoring method was used.

On the one hand, previous studies also suggest that switching to ideal point scoring could rarely yield significant improvements in criterion-related validity (Chernyshenko et al., 2007). As noted earlier, GGUM and CTT disagree about the rank-ordering of respondents among those relatively extreme respondents located beyond the point where folding occurs, and correlations will be insensitive to these (typically) minor and localized disagreements in rank-ordering. On the other hand, the negative impact of sum scores on criterion-related validity is likely to become more meaningful if the proportion of strongly folding items increases, if extreme respondents are the focus of prediction, or if the relationship with the criterion variable is nonlinear (Carter et al., 2014; Dalal & Carter, 2015).

Conclusions

In this research we compared a dominance model (GPCM) and an ideal point model (GGUM) of the process of responding to items from the CSIV. The results suggested that—although the CSIV was developed using CTT procedures that assumed a dominance response process—most CSIV items follow an ideal point response process. Other IPC inventories were developed using similar assumptions and procedures; however, it seems plausible that many of their items will also show an ideal point response process, and it seems prudent to test that possibility in future research.

Misspecification of an ideal point response process as a dominance response process might reduce the utility of CSIV. Inaccurate placements of individuals along latent trait dimensions could compromise diagnostic, selection, and classification decisions; in the case of the CSIV, the negative consequences of assuming a dominance response process might be particularly evident among respondents at the upper end of the trait dimensions, and particularly in the low-agency and low-communion octants of the IPC. More generally, misspecification of the item response process can prevent researchers from taking full advantage of the distinctive features of IRT described in the introduction (Stark et al., 2006).

That said, we do recognize the appeal of classical methods of estimating traits simply by aggregating items. IRT methods are more conceptually and mathematically complicated; consequently, using IRT methods might require recruiting larger samples, making stronger psychometric assumptions, and employing less familiar statistical programs than using traditional CTT-based sum or mean scores (Zickar & Broadfoot, 2009). Moreover, scores derived from CTT methods and scores derived from IRT methods often yield almost identical results. Thus, depending on the circumstances, the simpler CTT approach could sometimes be adequate.

On the other hand, the availability of free, user-friendly programs (e.g., the Windows-based GGUM2004 used in the current research) is reducing the barriers to using ideal point IRT scoring, making even relatively minor improvements in accuracy and utility worth the extra effort. Therefore, rather than estimating a respondent's trait levels for each CSIV octant in the traditional manner (by summing or averaging the responses to the items from each octant scale), in the future test users instead might consider estimating trait levels based on ideal point process assumptions. Moreover, to the degree that items linked to diverse personality constructs and response scales continue to show better fit to ideal point models than dominance response models, test developers might wish to give ideal point response models a greater role in guiding the development of new IPC measures.

Funding

This research was supported in part by grants from Guangzhou education scientific planning subject (Grant No. 1201411413) and the National Natural Science Foundation of China (Grant No. 31470050).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716−723.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Counseling and Clinical Psychology*, *42*, 155−162.

Carter, N. T., & Dalal, D. K. (2010). An ideal point account of the JDI Work Satisfaction scale. *Personality and Individual Differences*, *49*, 743−748.

Carter, N. T., Dalal, D. K., Boyce, A. S., O'Connell, M. S., Kung, M. C., & Delgado, K. M. (2014). Uncovering curvilinear relationships between conscientiousness and job performance: How theoretically appropriate measurement makes an empirical difference. *Journal of Applied Psychology*, *99*, 564−586.

Chernyshenko, O., Stark, S., Drasgow, F., & Roberts, B. (2007). Constructing personality scales under the assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. *Psychological Assessment*, *19*, 88−106.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Auflage). Hillsdale, NJ: Erlbaum.

Dalal, D. K., & Carter, N. T. (2015). Consequences of ignoring ideal point items for applied decisions and criterion-related validity estimates. *Journal of Business and Psychology*, *30*, 483−498.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Glas, C. A. (1988). The derivation of some tests for the Rasch model from the multinomial distribution. *Psychometrika*, *53*, 525−546.

Grassi, M., Luccio, R., & Di Blas, L. (2010). CircE: An R implementation of Browne's circular stochastic process model. *Behavior Research Methods*, *42*, 55−73.

Gurtman, M. B., & Pincus, A. L. (2003). The circumplex model: Methods and research applications. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (Vol. 2, pp. 407−428). New York, NY: Wiley.

Herrmann, A., & Pfister, H. R. (2013). Simple measures and complex structures: Is it worth employing a more complex model of personality in Big Five inventories? *Journal of Research in Personality*, *47*, 599−608.

Hopwood, C. J., Ansell, E. B., Pincus, A. L., Wright, A. G. C., Lukowitsky, M. R., & Roche, M. J. (2011). The circumplex structure of interpersonal sensitivities. *Journal of Personality*, *79*, 707−740.

Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, *14*, 332−346.

Kammrath, L. K., & Scholer, A. A. (2011). The Pollyanna myth: How highly agreeable people judge positive and negative relational acts. *Personality and Social Psychology Bulletin*, *37*, 1172−1184.

Locke, K. D. (2000). Circumplex scales of interpersonal values: Reliability, validity, and applicability to interpersonal problems and personality disorders. *Journal of Personality Assessment*, *75*, 249−267.

Locke, K. D. (2003). Status and solidarity in social comparison: Agentic and communal values and vertical and horizontal directions. *Journal of Personality and Social Psychology*, *84*, 619.

Locke, K. D. (2011). Circumplex measures of interpersonal constructs. In L. M. Horowitz & S. Strack (Eds.), *Handbook of interpersonal psychology: Theory, research, assessment and therapeutic interventions* (pp. 313−324). Hoboken, NJ: Wiley.

Locke, K. D., & Adamic, E. J. (2012). Interpersonal circumplex vector length and interpersonal decision making. *Personality and Individual Differences*, *53*, 764−769.

Locke, K. D., & Christensen, L. (2007). Re-construing the relational-interdependent self-construal and its relationship with self-consistency. *Journal of Research in Personality*, *41*, 389−402.

Locke, K. D., Craig, T., Baik, K. D., & Gohil, K. (2012). Binds and bounds of communion: Effects of interpersonal values on assumed similarity of self and others. *Journal of Personality and Social Psychology*, *103*, 879−897.

Locke, K. D., & Sadler, P. (2007). Self-efficacy, values, and complementarity in dyadic interactions: Integrating interpersonal and social-cognitive theory. *Personality and Social Psychology Bulletin*, *33*, 94−109.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*, 159−176.

Muraki, E., & Bock, D. (2003). Parscale for Windows (Version 4.1). Chicago, IL: Scientific Software International.

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, *4*, 207−230.

Roberts, J., Donoghue, J., & Laughlin, J. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement*, *24*, 3−32.

Roberts, J., Fang, H., Cui, W., & Wang, Y. (2006). GGUM2004: A Windows-based program to estimate parameters of the generalized graded unfolding model. *Applied Psychological Measurement*, *30*, 64−65.

Roberts, J., Laughlin, J., & Wedell, D. (1999). Validity issues in the Likert and Thurstone approaches to attitude measurement. *Educational and Psychological Measurement*, *59*, 211−233.

Roberts, J., & Shim, H. (2008). *GGUM2004 technical reference manual*. Atlanta, GA: Georgia Institute of Technology. Retrieved from http://www.psychology.gatech.edu.unfolding/FreeSoftware.html

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461−464.

Sodano, S. M., & Tracey, T. J. G. (2011). A brief inventory of interpersonal problems−circumplex using nonparametric item response theory: Introducing the IIP−C−IRT. *Journal of Personality Assessment*, *93*, 62−75.

Stark, S. (2001). *MODFIT: A computer program for model-data fit*. Champaign, IL: University of Illinois at Urbana−Champaign.

Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, *91*, 25−39.

Thomas, A., Kirchmann, H., Suess, H., Brautigam, S., & Strauss, B. M. (2012). Motivational determinants of interpersonal distress: How interpersonal goals are related to interpersonal problems. *Psychotherapy Research*, *22*, 489−501.

Turan, B., Guo, J., Boggiano, M. M., & Bedgood, D. (2014). Dominant, cold, avoidant, and lonely: Basal testosterone as a biological marker for an interpersonal style. *Journal of Research in Personality*, *50*, 84−89.

Turan, B., & Horowitz, L. M. (2010). The motive for support and the identification of responsive partners. *Journal of Research in Personality*, *44*, 342−352.

Van den Wollenberg, A. L. (1982). Two new test statistics for the Rasch model. *Psychometrika*, *47*, 123−140.

Wetzel, E., & Hell, B. (2014). Multidimensional item response theory models in vocational interest measurement: An illustration using the AIST−R. *Journal of Psychoeducational Assessment*, *32*, 342−355.

Wiggins, J. S. (2003). *Paradigms of personality assessment*. New York, NY: Guilford.

Zickar, M. J., & Broadfoot, A. A. (2009). The partial revival of a dead horse? Comparing classical test theory and item response theory. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity, and fable in the organizational and social sciences* (pp. 37−60). New York, NY: Routledge.