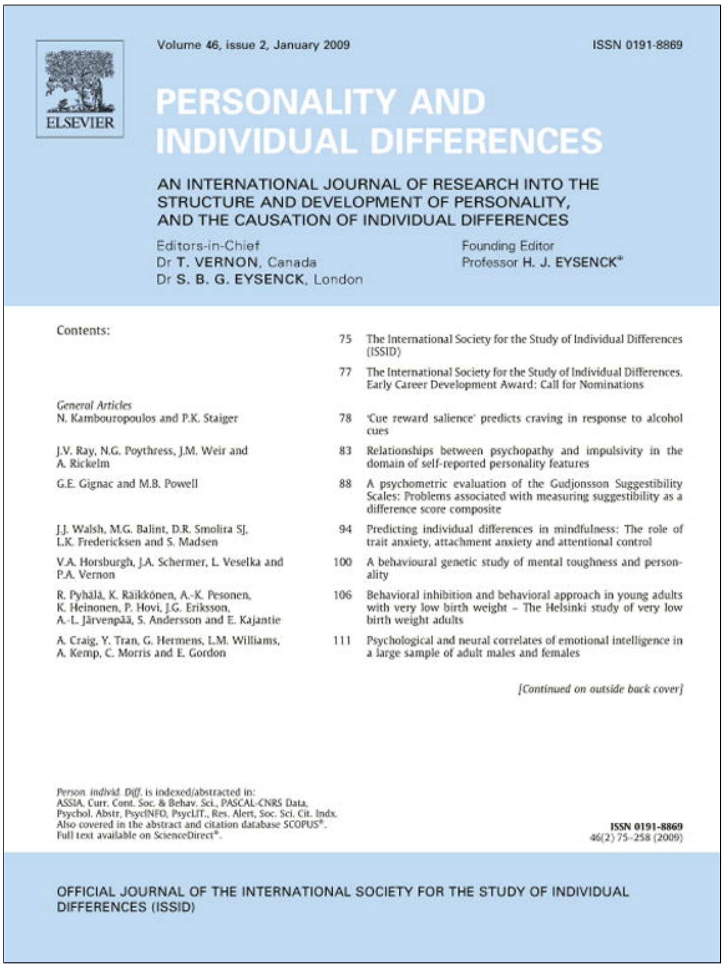

Volume 46, issue 2, January 2009

ISSN 0191-8869

ELSEVIER

PERSONALITY AND INDIVIDUAL DIFFERENCES

AN INTERNATIONAL JOURNAL OF RESEARCH INTO THE STRUCTURE AND DEVELOPMENT OF PERSONALITY, AND THE CAUSATION OF INDIVIDUAL DIFFERENCES

Contents:

[Continued on outside back cover]

Person. individ. Diff. is indexed/abstracted in:
ASSIA, Curr. Cont. Soc. & Behav. Sci., PASCAL-CNRS Data, Psychol. Abstr., PsycINFO, PsycLIT., Res. Alert, Soc. Sci. Cit. Indx.
Also covered in the abstract and citation database SCOPUS®.
Full text available on ScienceDirect®.

ISSN 0191-8869
46(2) 75–258 (2009)

OFFICIAL JOURNAL OF THE INTERNATIONAL SOCIETY FOR THE STUDY OF INDIVIDUAL DIFFERENCES (ISSID)

Personality and Individual Differences 46 (2009) 116–122

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Instructional set and the structure of responses to rating scales

Robert E. McGrath [a,*], Jessica Neubauer [a], Gregory J. Meyer [b], Kane Tung [a]

[a] School of Psychology T-WH1-01, Fairleigh Dickinson University, Teaneck NJ 07666, United States
[b] Department of Psychology, University of Toledo, 2801 West Bancroft Street, Toledo OH 43606, United States

A R T I C L E I N F O

Article history:
Received 6 June 2008
Received in revised form 8 September 2008
Accepted 16 September 2008
Available online 31 October 2008

Keywords:
Taxometric analysis
Rating scales
Self-report
Response bias
Instructional set

A B S T R A C T

The extent to which psychosocial phenomena are categorical or dimensional has been a topic of interest in recent years, in part spurred by the development of taxometric analysis as a statistical procedure for evaluating underlying structure (Meehl & Yonce, 1994, 1996; Waller & Meehl, 1998). Beauchaine & Waters (2003) suggested that the structure of scores on rating scales may be susceptible to instructional manipulation. However, they only investigated circumstances in which the ratings targets were unknown to the raters. The current study examined ratings of a small set of familiar targets. A sample of 608 undergraduate students completed five rating scales describing themselves and a significant other. Students were randomly assigned to instructional sets encouraging either dimensional or categorical ratings. Results consistently indicated a dimensional structure across both instructional sets and targets. The findings suggest that under normal conditions for the use of rating scales, instructional set does not affect the data structure.

1. Introduction

Research on whether psychosocial attributes are inherently dimensional or categorical entities has become popular in recent years. This interest can be traced to Paul Meehl's (e.g., Meehl, 1999) assertion that categorical or dimensional structure is an inherent characteristic of psychological constructs, and therefore can be evaluated empirically. Meehl and his colleagues developed a set of statistical methods, referred to collectively as taxometric analysis (Meehl & Yonce, 1994; Meehl & Yonce, 1996; Waller & Meehl, 1998), specific to addressing such questions. In response to Meehl's work, researchers have been looking more closely at the degree to which a categorical diagnostic model accurately portrays the character of psychological disorders (e.g., Haslam, 2007; Widiger & Trull, 2007) and the degree to which personality dimensions measured via psychometric measures reflect underlying categorical structure (e.g., Haslam & Kim, 2002).

Beauchaine and Waters (2003) conducted a study that raised concerns about the extent to which outcomes from taxometric analyses can potentially reflect situational factors rather than the inherent character of a psychological construct. In particular, they noted that many taxometric studies have relied upon data from rating scales, which are traditionally thought to be sensitive to response biases and other sources of systematic nuisance variability. To evaluate whether the structural quality of ratings is also sensitive to manipulation, they asked a sample of college students to rate essays submitted by 75 graduate school applicants. Raters were randomly told either that the essay writers' subsequent academic performance fell at the extremes of the distribution or reflected the full range of academic performance. The authors found raters given instructions suggesting extreme performance generated ratings that demonstrated a categorical structure, while raters given the instructions suggesting the essays reflected the full range of outcomes generated dimensional ratings. They concluded that researchers should "avoid an exclusive reliance on rating scales when conducting taxometrics investigations" (p. 3, and repeated on p. 10). If minor modifications in instructional set can significantly influence the structure of the resulting data, the implication is that structure can easily be manipulated by situational factors. This conclusion raises serious questions about the degree to which empirical analysis can detect true structure, particularly in cases where data is from rating scales.

Generalization of these findings to rating scales in general may be premature given the design of the study, however. The raters had no personal familiarity with the 75 targets of their ratings. This contrasts with the common use of rating scales to gather self-ratings or ratings by significant others. It is also noteworthy that in studies where the targets are known to the raters, it is common for raters to be rating no more than one, or perhaps a few, of the targets, reducing the opportunity to self-correct if the rater believes his or her ratings are not compliant with the instructional set. Finally, Beauchaine and Waters (2003) provided a financial incentive for the rater who could best predict which graduate students were successful to enhance compliance with

* Corresponding author. Tel.: +1 201 692 2445; fax: +1 201 692 2304.
E-mail address: mcgrath@fdu.edu (R.E. McGrath).

0191-8869/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.paid.2008.09.012

the instructions. The present study was conducted to evaluate whether instructional set would influence structure under more typical circumstances where raters evaluate a small set of targets known to them, without such a strong incentive to manipulate the structure of their responses.

To summarize, a study was conducted in which respondents completed rating scales under typical conditions, where the targets of the ratings are known to the respondent and there are no financial incentives for accuracy. One factor was manipulated, that being whether respondents were given an instructional set supportive of dimensional or categorical structure. If the results were to suggest that structure varied as a function of instructional set, they would raise serious concerns about the potential for the detection of true structure whenever rating scales are used. However, if the statistical results were not affected by instructions, this finding would support the use of rating scales for evaluation of known individuals as a basis for taxometric research on dimensional versus categorical structure.

2. Method

2.1. Participants

The initial sample consisted of 635 undergraduates from Fairleigh Dickinson University (NJ) and the University of Toledo (OH) who either volunteered or participated as part of a course research requirement during the period 2004–2005. After eliminating students who omitted more than two items on any one scale, the final sample included 608 students between the ages of 16 and 48.

2.2. Measures

The study involved five rating scales. Taxometric analyses require multiple measurements of a common core construct. Research demonstrates the five scales used in this study, while not interchangeable, overlap extensively as indicators of a single latent variable that has been referred to as core self-evaluation (Johnson, Rosen, & Levy, 2008; Judge, Erez, Bono, & Thoresen, 2002; Judge, Erez, Bono, & Thoresen, 2003; Judge, Locke, Durham, & Kluger, 1998), that is, the degree to which individuals evaluate themselves in a primarily positive or negative way. The Neuroticism Scale from the Eysenck Personality Inventory (Eysenck & Eysenck, 1968) consists of 12 items completed on a scale from 1 to 5. The Internality Subscale from Levenson's (1981) Internal, Powerful Others, and Chance Scale served as a measure of locus of control. It consists of eight items completed on a scale from 1 to 7. The Core Self-Evaluations Scale (Judge et al., 2003) was developed as a direct measure of the latent variable thought to underlie all five measures. It consists of 12 items completed on a scale from 1 to 5. The Generalized Self-Efficacy Scale (Judge et al., 1998) consists of eight items having to do with personal effectiveness completed on a scale from 1 to 7. The Rosenberg (1965) Self-Esteem Scale consists of 10 items completed on a scale from 1 to 4. In all cases 1 was anchored with the term *strongly disagree* while the highest option was anchored with *strongly agree*. Missing responses were replaced by the mean of the remaining items.

2.3. Procedure

All participants completed the set of questionnaires twice, once describing themselves and once describing a person they knew well. Order of the two targets was randomly counterbalanced. For the other-rating, item wording was changed to reference another person rather than self. All scales are usually keyed so that higher scores are indicative of more positive evaluations except the Neuroticism scale. This scale was key-reversed so that positive correlations were expected in all cases.

Raters were also randomly assigned to one of two instructional sets. Instructions were exactly the same for the two conditions, except that in one group the general instructions and the instructions for each scale included the sentence "previous studies on these types of scales suggest that most people tend to fall into one of two groups; they either produce very high or very low scores." In the second condition, this sentence was replaced with "previous studies suggest that people produce a wide range of scores on these types of scales, from low to medium to high." On the demographic sheet, respondents indicated the length and depth of their relationship with the significant other they rated. No participants were informed of the purpose of the study until all data were collected.

Taxometric analysis proceeds by conducting numerous tests of the taxonic hypothesis. Meehl and his colleagues derived a variety of taxometric methods, four of which were used in this study: MAMBAC, MAXCOV, MAXEIG, and L-Mode (Grove, 2004; Grove & Meehl, 1993; Meehl & Yonce, 1994; Meehl & Yonce, 1996; Ruscio, Haslam, & Ruscio, 2006; Waller & Meehl, 1998). The methods have been discussed in detail in the references cited, so will only be described briefly here. MAMBAC (Mean Above Minus Below A Cut) involves two variables measuring the same latent variable. Cut scores are set at successive points on one variable called the input indicator. At each cut score, the mean score for a second variable (called the output indicator) is computed separately for cases above and below the cut, and the difference between the means is computed. These output indicator mean differences are graphed as a function of the cut score for the input indicator. For taxonic constructs, this graph with cut scores on the abscissa and mean differences on the ordinate should be hill-shaped; when the latent construct is dimensional, the same graph should tend towards a U shape. Each pair of dimensional variables can be evaluated twice, with each variable serving as the input indicator for one analysis and the output indicator for the other.

MAXCOV (Maximum Covariance) requires one input and two output indicators. Observations are ordered and divided into overlapping subgroups (windows) along the input indicator. By default, 50 windows were created and the cases included in adjoining windows overlapped 90%. That is, the first window contained the 51 cases with the lowest scores on the input indicator. The second window included the 46 from the first window with the highest scores on the input indicator, and the five individuals with the next highest input indicator scores. This process was repeated until 50 windows had been created. The covariance between the two output indicators is then computed within each window. For taxonic constructs, a graph with input windows on the abscissa and output covariances on the ordinate should resemble a hill; if the construct is dimensional, the graph should be relatively flat or saw-toothed.

MAXEIG (Maximum Eigenvalue) is an extension of MAXCOV for circumstances where more than two variables are available for use as output indicators. After dividing observations into overlapping windows on the input indicator as described for MAXCOV, the eigenvalue for the first principal component based on the output indicators is computed within each window. The plot of eigenvalues as a function of window should follow the pattern described for MAXCOV. MAXEIG is usually conducted one time with each indicator as the input indicator and all others serving as the output indicators.

MAXCOV and MAXEIG also allow estimation of the Bayesian posterior probability of membership in the taxon class for each observation in the data set assuming a taxon is present. The histogram of these probabilities provides yet another test of the taxonic hypothesis. A histogram in which the probabilities divide into two sets that cluster near 0 and 1 is considered supportive of the presence of two classes. If instead the histogram reveals a single cluster of probabilities, or if probabilities are distributed across the entire range from 0 to 1, the results are considered more consistent with dimensional structure.

For L-Mode (Latent Mode; Waller & Meehl, 1998), all available indicators can be treated as a single set. The indicators are factor analyzed, and the factor scores are generated for the first factor. Because of the reduction in measurement error via factor analysis, the density plot of the scores should be bimodal if the data are taxonic and unimodal if the data are dimensional.

Taxometric analyses were conducted using R source code (Ruscio, 2008) in conjunction with R version 2.6.0. For each of the four sets of questionnaires (rating self and other, under dimensional or categorical instructions, the software generated 20 MAMBAC and MAXSLOPE curves, 30 MAXCOV curves, 5 MAXEIG curves, and 1 L-Mode curve.

Source code default options were used with three exceptions. Because of a technical issue with the software, the number of *X*-axis windows was increased from the default of 50–200 only for Bayesian classification based on MAXEIG. Second, tied scores on the input indicator can result in arbitrary placement of cases in one window or another. To reduce any effect attributable to this factor, each analysis was replicated 10 times to avoid any systematic bias in case classification. Finally, 10 simulated comparison data sets were generated under each structural model to allow comparison of data-based curves to expected taxonic and dimensional curve shapes. In addition, finding that the taxonic and dimensional models tend to produce very similar simulated distributions can suggest that a certain taxometric method may not be useful in a particular instance.

Table 1
Demographic statistics

	N	*M*	SD	%
Age	595	19.57	2.99	
Relationship length (years)	604	9.52	7.30	
Level of acquaintance[a]	606	3.80	0.47	
Location				
FDU	194			31.91
Toledo	414			68.09
Gender				
Male	221			36.53
Female	384			63.47
Year in college				
Freshman	378			62.58
Sophomore	140			23.18
Junior	52			8.61
Senior	34			5.63
Ethnicity				
White	374			61.92
Black	120			19.87
Hispanic	45			7.45
Asian	21			3.48
Other	44			7.28
Group				
Categorical	300			49.34
Dimensional	308			50.66
Relationship type				
Parent	51			9.24
Sibling	96			17.39
Friend	261			47.28
Other family	14			2.54
Romantic partner	130			23.55

[a] Reported on a scale from 1 (*not very well*) to 4 (*very well*).

3. Results

Demographic statistics may be found in Table 1. On average respondents knew the person they evaluated more than 9 years, a time frame that previous research would suggest is reasonable for establishing an accurate and stable perception of another person (e.g., Biesanz, West, & Millevoi, 2007; Paulhus & Bruce, 1992). Consistent with this assertion was the mean rating of 3.80 (SD = 0.47) on a scale indicating level of acquaintance from 1 (*not very well*) to 4 (*very well*).

Table 2 provides descriptive statistics for the five questionnaires under both the categorical and dimensional instructional sets. All correlations were significant except one ($p < .05$). As expected, within-sample correlations were generally large for both self-ratings ($M = .54$ for the categorical group, .61 for the dimensional group) and ratings of others ($M = .63$ for the categorical group, .62 for the dimensional group). Reliabilities were also generally acceptable, though reliabilities for the LOC scale were lower than the rest. Reliability could have been improved somewhat by eliminating one item having to do with the inevitability of future events. However, eliminating this item did not appreciably affect the size of correlations with the other scales, so the item was retained to maintain comparability with previous research.

Graphs relevant to evaluating the hypothesis of taxonic structure may be found in Figs. 1–4 and Table 3. Each figure provides graphic results for one of the four sets of questionnaires. For MAMBAC, MAXCOV, and MAXEIG, the figures present the average of data-based curves. Inspection of the individual curves indicates there was very little deviation from the average. The left curve in the graph is superimposed on an expected range for the curve if the data were derived from a structurally taxonic latent variable. The curve on the right reflects the expected range if the data were based on a dimensional structure. These expected ranges were derived from the simulated data sets. The same comparison is provided for L-Mode, though the results are based on a single data-based curve. The histogram of probabilities for Bayesian classification analyses is also provided.

Table 3 provides results from three statistics relevant to the evaluation of structure. Waller and Meehl (1998) introduced a goodness of fit index for the taxonic model. A value for this statistic ⩾.90 is thought to be indicative of taxonic structure, while values <.90 suggest dimensional structure. The software used in this study also allowed computation of a comparison curve fit index (CCFI). This statistic evaluates consistency between the data-based curves and the simulation data generated using the assumption of taxonic structure and then again under the assumption of dimensional structure. These simulated data sets are the same used to generate the boundaries for the expected ranges in Figs. 1–4. For the CCFI, values >.50 suggest the actual curves are more consistent with the simulated taxonic data, values <.50 suggest greater similarity with dimensional simulations, and values close to .50 are neutral.

Finally, each curve allows estimation of the taxon base rate assuming a taxon exists. Though L-Mode only produces one curve, three taxon base rate estimates can be generated using the two most frequently occurring values in the distribution and the classification cases. The standard deviation of these three estimates was computed. Values <.10 are thought to be consistent enough to suggest taxonic structure.

It should be noted that serious questions have been raised about the accuracy of the goodness of fit, variability, and Bayesian approaches to evaluating structure (e.g., Cleland, Rothschild, & Haslam, 2000; Ruscio, 2007; Ruscio, Ruscio, & Meron, 2007; Ruscio et al., 2006). The CCFI has also been questioned by Beach, Amir, and Bau (2005), though a recent reanalysis of their data supported its use in taxometric research (Ruscio & Marcus, 2007). Given that some controversy surrounds each of the techniques developed for evaluating taxonic structure except the primary graphs, the results of these additional analyses should only be used if they demonstrate a fairly high degree of consistency in interpretation.

Table 2
Correlations, means, and standard deviations for the scales

	Dimensional											
	EPI-S	CSE-S	LOC-S	GSE-S	RSE-S	EPI-O	CSE-O	LOC-O	GSE-O	RSE-O	*M*	*SD*
Categorical												
EPI-S	**0.88**	0.65	0.30	0.53	0.57	0.37	0.26	0.09	0.26	0.22	50.43	9.18
CSE-S	0.58	**0.82**	0.57	0.79	0.76	0.26	0.29	0.17	0.35	0.29	44.48	6.99
LOC-S	0.19	0.52	**0.58**	0.58	0.50	0.11	0.19	0.35	0.27	0.19	39.38	5.57
GSE-S	0.47	0.71	0.52	**0.84**	0.80	0.17	0.27	0.19	0.43	0.29	45.08	8.04
RSE-S	0.52	0.77	0.44	0.73	**0.88**	0.20	0.29	0.13	0.35	0.33	33.54	5.72
EPI-O	0.36	0.33	0.14	0.32	0.30	**0.89**	0.65	0.29	0.57	0.58	51.90	10.17
CSE-O	0.36	0.47	0.26	0.39	0.39	0.70	**0.87**	0.58	0.80	0.80	44.39	8.73
LOC-O	0.15	0.21	0.34	0.21	0.15	0.33	0.57	**0.65**	0.62	0.52	39.31	6.54
GSE-O	0.30	0.44	0.30	0.50	0.40	0.55	0.80	0.63	**0.87**	0.83	44.43	8.32
RSE-O	0.28	0.41	0.22	0.35	0.42	0.55	0.78	0.55	0.83	**0.90**	33.18	6.20
M	51.32	45.60	40.33	46.57	34.34	53.25	45.00	39.01	45.01	33.54		
SD	10.13	7.12	5.29	6.65	4.99	10.39	8.40	6.30	7.99	5.83		

Note. Bolded diagonal values are reliabilities for the sample as a whole. Values above the diagonal are descriptive statistics for the sample given dimensional instructions ($N = 308$). Values below the diagonal are descriptive statistics for the sample given categorical instructions ($N = 300$). EPI = Eysenck Personality Inventory Neuroticism; CSE = Core Self-Evaluation; LOC = Locus of Control; GSE = Generalized Self-Efficacy; RSE = Rosenberg Self-Esteem; Self = self-rating; O = other rating. Italicized correlations are not significant ($p < .05$).

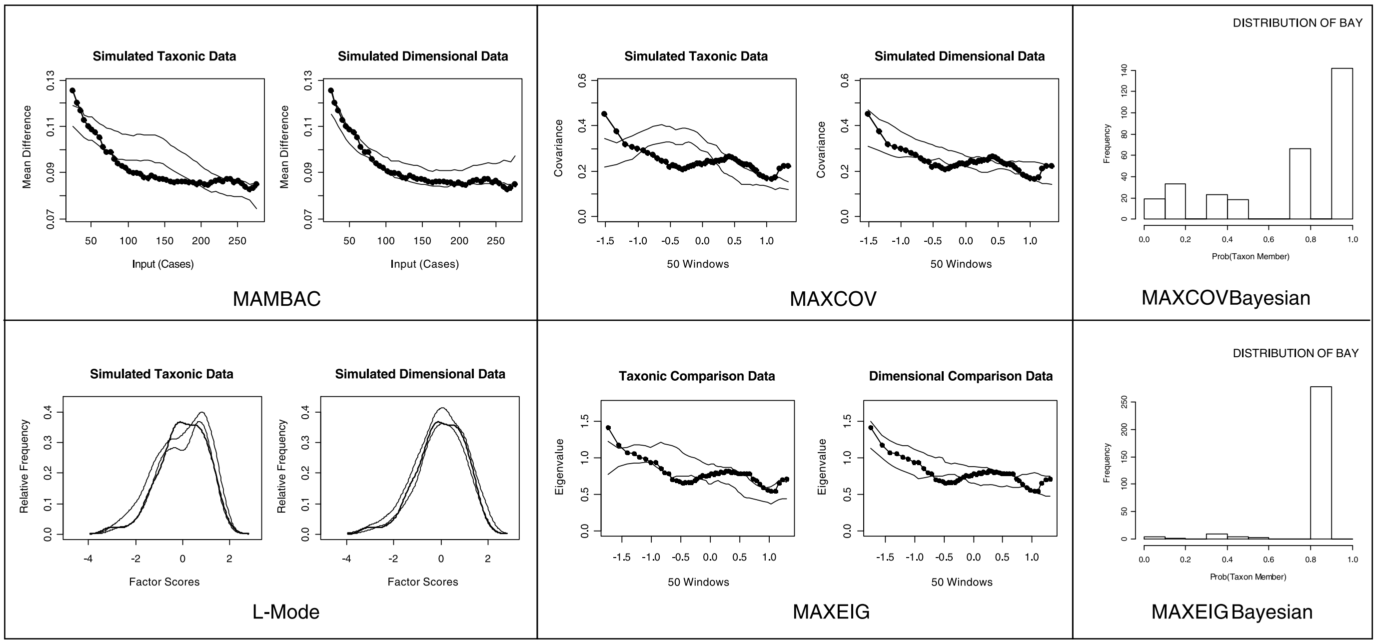

Fig. 1. Taxometric curve results for self-ratings based on categorical instructions ($N = 300$).

In all four figures, the results are more consistent with dimensional than taxonic structure. This is perhaps most easily seen in the Bayesian histograms. Taxonic structure should tend to produce two bars at the extremes of the graph. Instead, the results consistently suggested a series of clusters across the entire range of Bayesian probabilities. The MAMBAC curves were more difficult to interpret, with relatively little difference between the simulated taxonic and dimensional data models. However, the data-based curves tended to comply more with expectations based on the latter model. The same was true for the MAXEIG and MAXCOV data-based curves. Finally, L-mode curves were consistently unimodal.

All three statistics also consistently suggested dimensional structure. All goodness of fit indices were below .90. Every CCFI either suggested a dimensional conclusion or was neutral (between .40 and .60). Finally, the pattern of standard deviations was the same regardless of instructional set. Across every method used, the results were Ambiguous or suggested the structure underlying the data was dimensional.

4. Discussion

Beauchaine and Waters (2003) provided a clear basis for concern about the use of taxometric analysis to detect latent structure when the data are derived from rating scales. However, their research design reflected several elements that raise doubts about the generalizability of their findings to the standard circumstances under which rating scales are administered. First, respondents rated numerous individuals unknown to them. Second, respondents were given a financial incentive to comply with the instructions. When both of these factors were removed, the results provide consistent evidence of dimensional structure regardless of whether the instructional set implied a dimensional or categorical structure, and whether the target was the respondent or a significant other. Instructional set seems to have played very little role in determining how participants responded to the questionnaires. These results also offer evidence to suggest that core self-evaluation is an inherently dimensional construct. That is, individ-

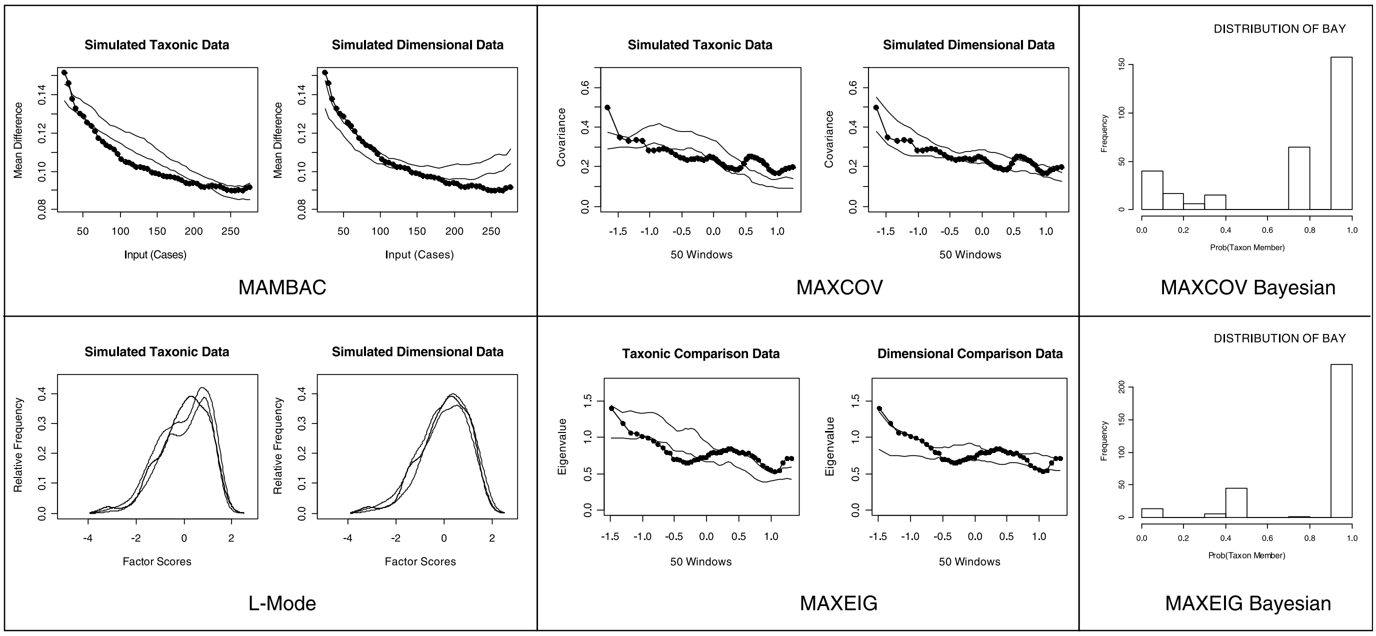

Fig. 2. Taxometric curve results for other-ratings based on categorical instructions ($N = 300$).

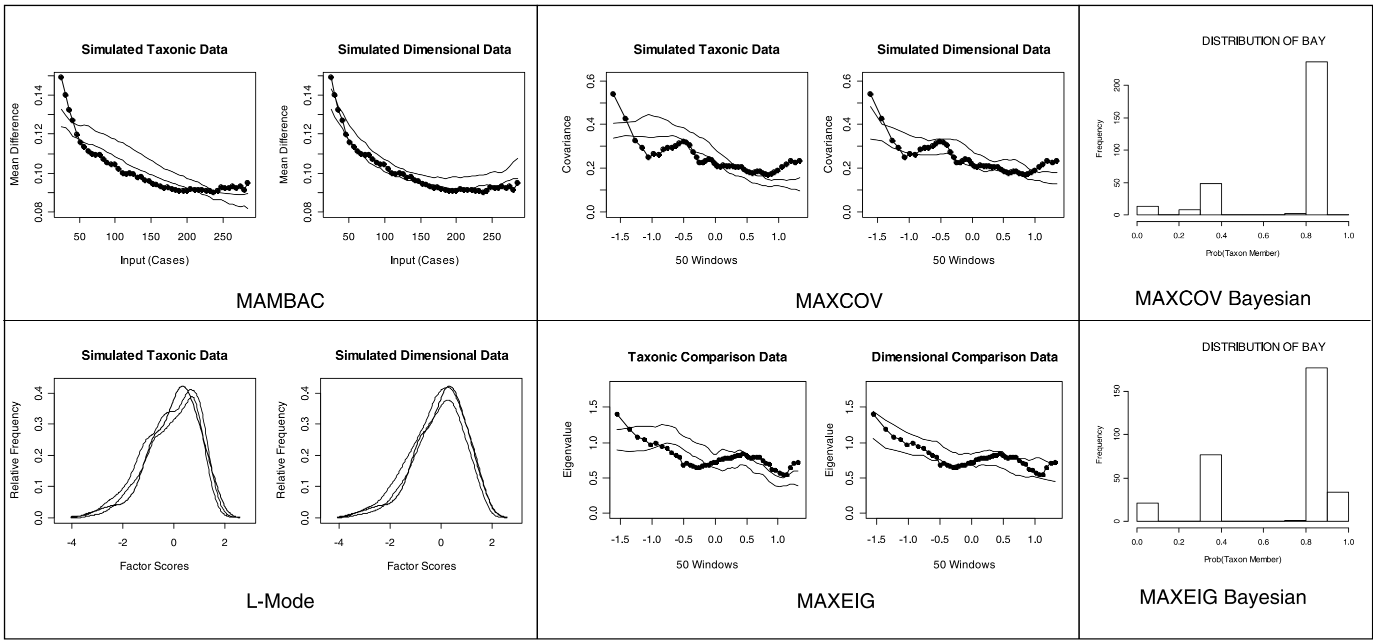

Fig. 3. Taxometric curve results for self-ratings based on dimensional instructions ($N = 308$).

uals evaluate themselves and significant others along a dimension from positive to negative.

One possible limitation of this study has to do with the effectiveness of the instructional intervention, whether the participants even attended to the structural bias. No manipulation check was included to avoid hypothesis guessing. However, it is noteworthy that the participants in Beauchaine and Waters' (2003) study received the instructions only once. In the current study the instructions were incorporated into the overall instructions for the study and in the instructions for each questionnaire, for a total of 11 presentations. The failure to find modification of structure even in the face of explicit repetition of the expected structure suggests respondent behavior is unlikely to be modified except in rather unusual circumstances such as those studied by Beauchaine and Waters.

Given that the central issue in this study has to do with the generalizability of a finding, it is worth noting that both the Beauchaine and Waters (2003) and the current study focused on a single construct. Core self-evaluation was chosen for this study in part because it encompasses a set of scales that reflect constructs commonly evaluated in rating scale research (self-esteem, locus of control, etc.), but the possibility exists that other aspects of the constructs involved such as familiarity, or whether the construct is inherently categorical or dimensional, may have moderated the effectiveness of the instructional set.

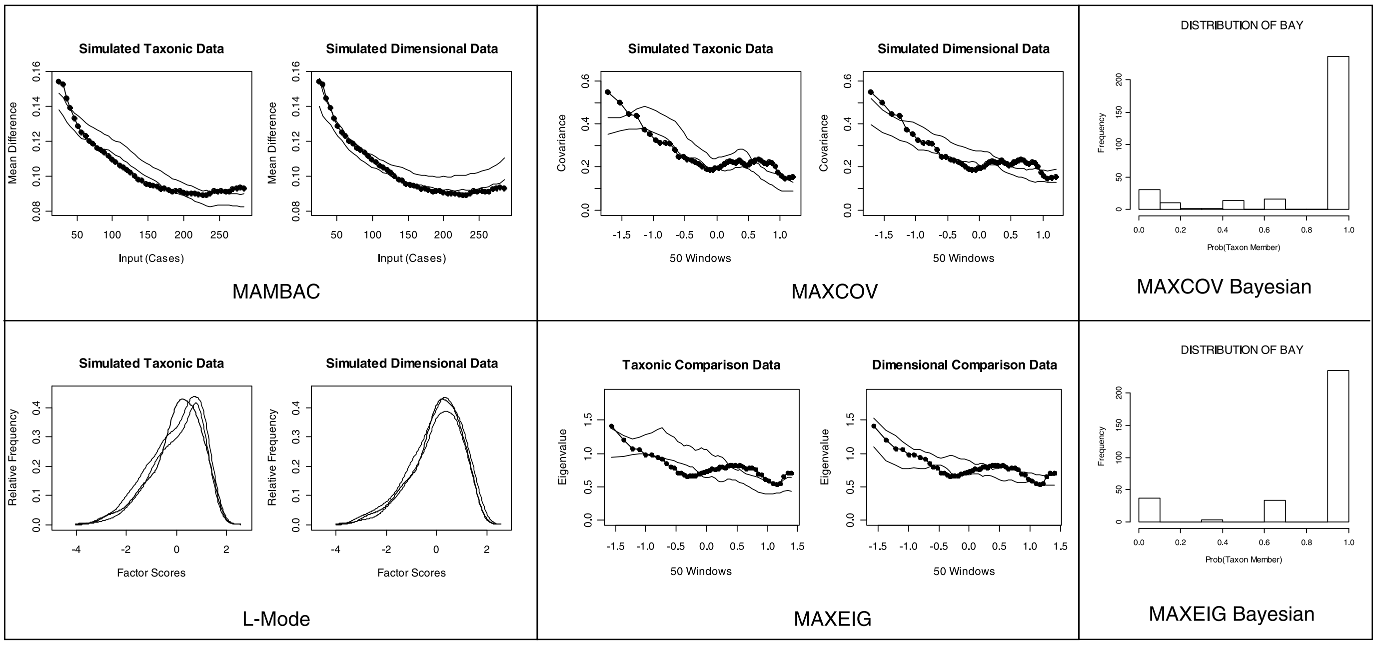

Fig. 4. Taxometric curve results for other-ratings based on dimensional instructions ($N = 308$).

Table 3
Goodness of fit indices for taxonic and dimensional structure

	Categorical self			Categorical other			Dimensional self			Dimensional other		
	GFI	CCFI	SD	GFI	CCFI	SD	GFI	CCFI	SD	GFI	CCFI	SD
MAMBAC	0.81	0.42	0.08	0.72	0.58	0.05	0.75	0.32	0.06	0.71	0.48	0.08
MAXCOV	0.76	0.33	0.23	0.69	0.31	0.18	0.41	0.37	0.22	0.71	0.48	0.12
MAXEIG	0.77	0.39	0.04	0.77	0.38	0.04	0.77	0.33	0.04	0.77	0.35	0.04
L-Mode	0.81	0.37	0.49	0.70	0.35	0.46	0.73	0.31	0.45	0.68	0.33	0.48

Note. GFI = goodness of fit index; CCFI = comparison curve fit index; SD = standard deviation of base rate estimates across curves.

It is worth reiterating that the present study differed from the earlier one on three relevant dimensions, to more closely match circumstances common in the use of rating scales. The first was the rating of individuals known to the respondent. The second was that respondents only rated two targets. The third was the absence of a financial incentive to follow either instructional set. It is unclear to what extent each was important to the earlier study's findings. Given prior evidence that familiarity is an important moderator of responses to rating scales (e.g., Biesanz et al., 2007; Norman & Goldberg, 1966), it is reasonable to assume this factor was an important contributor to the difference. However, it seems likely that all three design features combined to facilitate Beauchaine and Waters' findings.

The results support the continued use of taxometric methods to evaluate structure in rating scales when respondents are evaluating individuals familiar to them. This is particularly likely to be true when rating scales are used for self-report, since each respondent by definition is rating a single individual familiar to them, or whenever the number of targets per respondent is limited. Under these circumstances, familiarity with the target seems to trump instructional set or prior expectations about structure. An issue that has yet to be adequately explored is the extent to which self- or other-ratings can be manipulated to affirm dimensional or categorical structure by secondary gains. This is likely to be a salient concern when multiple targets are being rated by a single judge.

Acknowledgements

We are grateful to Nicholas Katko, Konstantin Lukin, and Matthew Mitchell for their help creating the data set for this study, to Theodore Beauchaine and Timothy Judge for providing materials from prior studies that were adapted for the present investigation, and to John Ruscio for guidance in the use of the taxometric software.

References

Beach, S. R. H., Amir, N., & Bau, J. J. (2005). Can sample-specific simulations help detect low base-rate taxonicity? *Psychological Assessment, 17*, 446–461.
Beauchaine, T. P., & Waters, E. (2003). Pseudotaxonicity in MAMBAC and MAXCOV analyses of rating scale data: Turning continua into classes by manipulating observer's expectations. *Psychological Methods, 7*, 245–261.
Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology, 92*, 119–135.
Cleland, C. M., Rothschild, L., & Haslam, N. (2000). Detecting latent taxa: Monte Carlo comparison of taxometric, mixture model, and clustering procedures. *Psychological Reports, 87*, 37–47.
Eysenck, H. J., & Eysenck, S. B. G. (1968). *Manual of the Eysenck personality inventory*. San Diego, CA: Educational and Industrial Testing Service.
Grove, W. M. (2004). The MAXSLOPE taxometric procedure: Mathematical derivation, parameter estimation, consistency tests. *Psychological Reports, 95*, 517–550.
Grove, W. M., & Meehl, P. E. (1993). Simple regression-based procedures for taxometric investigations. *Psychological Reports, 73*, 707–737.

Haslam, N. (2007). The latent structure of mental disorders: A taxometric update on the categorical vs. dimensional debate. *Current Psychiatry Reviews, 3*, 172–177.

Haslam, N., & Kim, H. C. (2002). Categories and continua: A review of taxometric research. *Genetic, Social, and General Psychology Monographs, 128*, 271–320.

Johnson, R. E., Rosen, C. C., & Levy, P. E. (2008). Getting to the core of self-evaluation: A review and recommendations. *Journal of Organizational Behavior, 29*, 391–413.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct? *Journal of Personality and Social Psychology, 83*, 693–710.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2003). The core self-evaluations scale: Development of a measure. *Personnel Psychology, 56*, 303–331.

Judge, T. A., Locke, E. A., Durham, C. C., & Kluger, A. N. (1998). Dispositional effects on job and life satisfaction: The role of core evaluations. *Journal of Applied Psychology, 83*, 17–34.

Levenson, H. (1981). Differentiating between internality, powerful others, and chance. In H. M. Lefcourt (Ed.). *Research with the locus of control construct* (vol. 1, pp. 15–63). New York: Academic Press.

Meehl, P. E. (1999). Clarifications about taxometric method. *Applied and Preventive Psychology, 8*, 165–174.

Meehl, P. E., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonicity using two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports, 74*, 1059–1274.

Meehl, P. E., & Yonce, L. J. (1996). Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV procedure). *Psychological Reports, 78*, 1091–1227.

Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology, 4*, 681–691.

Paulhus, D. L., & Bruce, M. N. (1992). The effect of acquaintanceship on the validity of personality impressions: A longitudinal study. *Journal of Personality and Social Psychology, 63*, 816–824.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Ruscio, J. (2007). Taxometric analysis: An empirically-grounded approach to implementing the method. *Criminal Justice and Behavior, 24*, 1588–1622.

Ruscio, J. (2008). *Taxometric programs for the R computing environment: User's manual*. Ewing, NJ: The College of New Jersey Psychology Department.

Ruscio, J., Haslam, N., & Ruscio, A. M. (2006). *Introduction to the taxometric method: A practical guide*. Mahwah, NJ: Lawrence Erlbaum.

Ruscio, J., & Marcus, D. K. (2007). Detecting small taxa using simulated comparison data: A reanalysis of beach, Amir, and Bau's (2005) data. *Psychological Assessment, 19*, 241–246.

Ruscio, J., Ruscio, A. M., & Meron, M. (2007). Applying the bootstrap to taxometric analysis: Generating empirical sampling distributions to help interpret results. *Multivariate Behavioral Research, 42*, 349–386.

Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua*. Thousand Oaks, CA: Sage.

Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist, 62*, 71–83.