Categories versus dimensions in personality and psychopathology:
A quantitative review of taxometric research

Nick Haslam       Elise Holland
University of Melbourne

Peter Kuppens
University of Leuven

in press, Psychological Medicine
Abstract

Taxometric research methods were developed by Paul Meehl and colleagues to test between categorical and dimensional models of latent variables. A comprehensive review of published taxometric research was conducted that included 177 articles, 311 distinct findings, and a combined sample of 533,377 participants. Multilevel logistic regression analyses examined the methodological and substantive variables associated with taxonic findings. Although 38.9% of findings were taxonic, these findings were much less frequent in more recent and methodologically stronger studies, and in those reporting comparative fit indices based on simulated comparison data. When these and other possible confounds were statistically controlled the true prevalence of taxonic findings was estimated at 14%. The domains of normal personality, mood disorders, anxiety disorders, eating disorders, externalizing disorders, and personality disorders other than schizotypal yielded little persuasive evidence of taxa. Promising but still not definitive evidence of psychological taxa was confined to the domains of schizotypy, substance use disorders, and autism. The review indicates that most latent variables of interest to personality and clinical psychologists are dimensional, and that many influential taxonic findings of early taxometric research are likely to be spurious.

Keywords: categories, dimensions, personality, psychopathology, taxometrics
Deciding between categorical and dimensional models of latent variables is a fundamental and enduring issue in psychiatry and psychology, which taxometric analyses were designed to resolve. Developed by Paul Meehl and his colleagues, these procedures allow researchers to determine whether observed variation is underpinned by a nonarbitrary latent class, or “taxon,” such as a discrete psychopathology or personality type. Discovering taxa and distinguishing them from latent dimensions has broad implications for how personality and psychopathology should be conceptualized, assessed, and explained.

If a latent variable is taxonic, for example, it must be conceptualized as an entity with real category boundaries that exist independent of social convention or descriptive convenience. If it is not taxonic then no boundary exists unless a manifest distinction such as a diagnostic threshold is imposed on arbitrary or pragmatic grounds. The appropriate way to assess a taxonic variable involves assigning cases to categories at the taxon boundary, but assessing nontaxonic variables involves quantifying variation along the entirety of an underlying continuum. Taxa are likely to spring from mechanisms that Meehl (1977) referred to as “specific etiologies,” such as single discrete causal factors, whereas nontaxonic variables generally result from the additive effects of multiple small causal influences. Determining whether or not a latent variable is best thought of as taxonic is a crucial scientific question and not merely a matter of theoretical taste or statistical botanizing.

The taxometric method makes this determination in a distinctive way. Unlike some more familiar statistical approaches to latent variable analysis it does not impose a particular kind of structure, as cluster analysis presumes a categorical structure or factor
analysis a set of underlying dimensions, but instead tests between these alternatives. Unlike most comparable forms of data analysis it does not follow a null hypothesis testing approach to inference or yield a single definitive statistic. Where most other analyses employ a single statistical procedure, the taxometric method seeks consistency among the findings of multiple mathematically independent procedures. Other analyses provide chiefly numerical output, whereas the output of taxometric analyses is largely graphical, based on the interpretation of curves. Despite these unusual features, the taxometric method has proven to be popular and versatile (Ruscio, Haslam, & Ruscio, 2006).

Taxometric analyses were first employed to test Meehl’s theory of schizophrenia, which proposed a taxonic genetic liability that manifests as schizotypal personality. The initial study (Golden & Meehl, 1979) supported the existence of the proposed taxon, and in the 1980s a series of articles reported taxonic findings in the domains of normal personality (self-monitoring: Gangestad & Snyder, 1985; Type A: Strube, 1989), abnormal personality (schizotypy: Erlenmeyer-Kimling, Golden, & Cornblatt, 1989), and other psychiatric phenomena (dementia: Golden, 1982; tardive dyskinesia: Golden, Campbell, & Perry, 1987; nuclear depression: Grove et al., 1987). The trickle of studies in the 1980s became a stream in the 1990s and a torrent in the new millennium.

The fundamental questions that motivated the taxometric method are still pressing. Within clinical psychology and psychiatry the categorical/dimensional issue remains contentious as many theorists, clinicians, and researchers question the merits of categorical diagnosis and classification. These doubts have reached a crescendo, playing a role in the changes underway in DSM-V (Helzer, Kraemer, & Krueger, 2008; Widiger
& Samuel, 2005), which embeds dimensional judgments in psychiatric diagnosis like never before, recognizing degrees of severity and impairment in many conditions and diagnosing personality disorders along a set of continua.

The structural question that drives taxometric research also has continuing relevance to the psychology of normal personality. Dimensional views of traits predominate (Meehl, 1992), but there is enduring interest in the possible existence of personality types (e.g., Asendorpf, 2002) or configural prototypes (Eaton, Krueger, South, Simms, & Clark, 2011). Personality psychology’s status quo is thus the mirror image of psychiatry’s: evidence for personality types challenges the default dimensional assumptions of trait psychologists, just as evidence that psychopathology is a matter of degree challenges categorical assumptions about psychiatric diagnosis. Taxometric research can provide a firmer empirical foundation for scientific taxonomy in both fields by testing these assumptions.

Taxometrics is not the only form of data analysis that can test between categorical and dimensional models of latent variables, and sophisticated alternatives exist (De Boeck, Wilson, & Acton, 2005; Markon & Krueger, 2006; McLachlan & Peel, 2000). Nevertheless, taxometric research is a particularly rich source of knowledge on latent structure within psychology and psychiatry because its longevity means that it has built up a relatively large body of empirical findings on a wide assortment of latent variables. Reviewing those findings could help to answer basic questions about the latent structure of psychological variation. First, does existing research support the existence of any taxa in personality and psychopathology? Second, if solid evidence of taxa has been obtained, in what domains are they found? Taxa may be rarer in normal personality than in
psychopathology, and rarer in personality disorders (PDs) than in other psychiatric domains (Trull & Durrett, 2005). Whether taxa differ in prevalence across these domains and whether there is robust and replicated evidence for particular taxa are open questions.

In addition to clarifying substantive questions regarding the latent structure of personality and psychopathology, a systematic review of taxometric research might also shed light on methodological issues. First, are certain sample types (e.g., “abnormal” clinical or forensic samples) more likely to yield taxonic findings than others (e.g., “normal” community members)? Similarly, are taxa less frequent among children and adolescents than among adults, as might be expected if taxa differentiate developmentally? Sample size may also have implications for taxonic findings, as research employing samples that fail to meet Meehl’s (1995) recommended minimum $N$ of 300 may have a taxonic bias because they generate less stable curves. Second, are certain kinds of data more or less likely to produce taxonic findings? Some writers have argued that self-report data can generate spurious taxonic findings (Beauchaine & Waters, 2003).

A third set of methodological questions relates to aspects of measurement in taxometric research, which relies on multiple “indicators” of the proposed latent variable. Researchers have argued that dichotomous indicators may induce a taxonic bias (Ruscio, 2000), and that inadequate indicator validity may also impair taxometric inference (Meehl & Yonce, 1994, 1996). Research that employs indicators of limited reliability (e.g., based on single questionnaire items) or fails to report evidence of validity may yield taxonic findings at different rates than research employing more valid indicators.
A final set of methodological questions involves statistical methods. Five taxometric procedures are in widespread use (L-Mode, MAMBAC, MAXCOV, MAXEIG, MAXSLOPE) and some may be more likely to draw taxonic conclusions. Finding consistency among multiple procedures is a hallmark of the taxometric method, and the use of relatively few procedures can be considered one element of low methodological quality, along with small sample sizes, and dichotomous, single-item, and unvalidated indicators, that might have a bearing on taxometric research findings.

One statistical method that deserves special mention is the use of simulated comparison data. This adjunct to taxometric practice (Ruscio, Ruscio, & Meron, 2007) involves the parallel analysis of simulated taxonic and dimensional data sets that match the distributional and correlational properties of the observed research data. If the graphical output generated from the research data more closely resembles the output of one set of simulations than the other, then the simulation procedure supports the corresponding latent structure. The comparative fit of the research data output to the two simulations can be quantified by an index (the comparison curve fit index: CCFI), where 0.5 represent equally good fit, values less than 0.5 support a dimensional finding and values greater than 0.5 support taxonicity. The CCFI offers an objective decision rule to supplement visual inspection of curves, the traditional basis for taxometric inference.

Although the simulated comparison data procedure has been controversial among some researchers (Beach, Amir, & Bau, 2005; cf. Ruscio & Marcus, 2007), Monte Carlo studies provide very strong evidence for its validity and robustness under unfavourable data conditions. In an analysis of 25,000 simulated data sets constructed to present challenges for taxometric inference (e.g., indicator skew and coarseness, modest indicator
validity, “nuisance covariance” among indicators, unequal variance of latent 
distributions), Ruscio and Kaczetow (2009) found that the CCFI achieved 93% accuracy 
in identifying taxonic and dimensional latent structures, rising to 98% when CCFI values 
were outside an ambiguous intermediate range (i.e., <0.4 or >.6). In an even larger study 
of 100,000 data sets, Ruscio et al. (2010) found that a CCFI threshold of 0.5 achieved an 
average of 94% accuracy for MAMBAC, MAXCOV, and L-Mode applied individually, 
and 98% when the mean CCFI of the three procedures was used. A similar study of 
10,000 data sets (Ruscio et al., 2007) found that the CCFI strongly outperformed several 
previous fit indices and consistency tests. Use of simulated comparison data and the 
CCFI has become widespread, and it is therefore important to review whether this 
methodological development has had an influence on taxometric research findings.

Previous reviews of taxometric research have been qualitative and restricted to 
psychopathology-related constructs (Haslam, 2003, 2007), and they are now seriously 
out-dated, the last comprehensive review (Haslam, 2002) covering less than one quarter 
of the taxometric articles published as of 2011. We therefore conducted a comprehensive 
quantitative review of all published empirical findings using multilevel logistic regression 
analysis, in an effort to ascertain the prevalence of taxonic findings and the factors 
associated with them. Although the primary focus of the study was on substantive factors 
(i.e., which latent variables are taxonic), we also examined methodological factors that 
might contribute to taxonic findings, potentially as sources of bias.
Method

Study sample

An exhaustive literature search was conducted using previous reviews, publication databases (Google Scholar, PsycInfo, Web of Science), and journal and publisher websites (search terms included taxometric*, taxon*, MAXCOV, MAMBAC, MAXEIG). Inclusion and exclusion criteria were as follows: research had to 1) be published in peer-reviewed journal articles; 2) be officially published or “in press” by April 1, 2011; 3) address observed substantive latent variables rather than simulated data sets or those involving an experimental manipulation (Arnau, Thompson, & Cook, 2001; Beauchaine & Waters, 2003; McGrath, Neubauer, Meyer, & Tung, 2009); and 4) employ at least one of the accepted taxometric procedures.

Each article could yield one or more pertinent empirical finding, defined as a conclusion about the latent structure of a single construct based on one or more taxometric procedures in a single sample. Different findings within one article could reflect the empirical investigation of more than one construct and/or be based on distinct samples. In this sense, when multiple taxometric procedures are used to analyse a single construct in a single sample they contribute to a single finding. Similarly, parallel analyses of different sets of indicators of a single construct in a single sample were counted as contributing to a single finding. On these definitions, the 177 articles contained 311 findings ($M=1.76$, range 1-11).
Coding

The 177 articles were coded by the first author on multiple characteristics under the headings of publication details, sample characteristics, measurement characteristics, data analysis, results, and construct. All coding was conducted at the level of the individual finding, except for codes related to publication details (i.e., publication year, journal, number of findings), which were conducted at the article level.

Sample characteristics. For each finding, we coded sample size, whether the sample was composed of undergraduates, whether it was drawn from a clinical or forensic sample, whether the sample was drawn from the general public, and whether it was comprised primarily of children or adolescents (<18 years).

Measurement characteristics. We coded whether any of the data were based on self-report, ratings by observers, or interviews. With respect to indicator construction we coded the number of indicators employed in the analysis, whether any indicators were based on a single item rather than summed items, whether any indicators were dichotomous, and whether the validity of the indicators was reported in the manuscript (including quantification of indicator validities or any mention that they had satisfied a quantitative validation process).

Data analysis. Data-analytic methodology was coded in terms of the use or non-use of the MAXCOV, MAMBAC, MAXEIG, MAXSLOPE, L-Mode, and “other” taxometric procedures. These codes were summed to produce a “number of procedures” variable. Use of Ruscio’s simulated comparison data technique was also coded.

Results. The overall conclusion for each finding was coded taxonic (1) or nontaxonic (0), based on the researchers’ interpretation. Two ambiguous interpretations
were coded as nontaxonic. For studies reporting CCFI values, or allowing them to be computed from an earlier fit index (fit_{RMSR}), we coded the values for each taxometric procedure that was used to generate them. For each finding with at least one CCFI reported, a mean CCFI value was computed by taking the average across procedures.

**Construct.** The studied constructs were classified into ten groupings, developed to reflect broad construct domains but also to recognize narrower domains that have received substantial taxometric attention. Seven groupings were psychopathology-related, one referred to normal personality, and two were residual groupings. The groupings and a listing of constructs examined in the article sample are presented in Table 1.

The “mood disorder” grouping included mood disorders, proposed subtypes of these disorders, affective phenomena related to them (grief, mania), and diatheses for them (depression-proneness). The “anxiety disorder” grouping similarly included a mix of disorders, proposed variants, specific kinds of anxiety or aversion, and relevant diatheses (anxiety and disgust sensitivity). The “eating disorder” grouping followed the same pattern. The “substance use” grouping included abuse and dependence on a variety of substances. The “externalizing” grouping contained constructs involving antisocial and under-controlled conduct, including gambling and sexual behavior. A “schizotypy” grouping was defined narrowly given the longstanding tradition of taxometric research on this topic, and was kept separate from an “other personality disorder” grouping which included studies of seven recognized or proposed PDs. The “normal personality” grouping comprised personality or temperamental traits. The “other individual difference” grouping contained diverse psychological constructs including response styles and biases, attitudes, and pathological phenomena that have received little
taxometric attention and do not readily fit under the other groupings. The final
“miscellaneous” grouping contained constructs that either were not psychological (e.g.,
biological sex, metabolic syndrome) or did not have the individual as the unit of analysis
(e.g., emotions, relationship types).

**General considerations**

Individuating distinct findings within studies was occasionally challenging, as
multiple parallel analyses were sometimes reported using distinct indicator sets and/or
different subsamples (e.g., men versus women). In these circumstance, if a single
construct was being examined we coded analyses based on the entire sample and not
those based on subsamples, and we took an average value of all quantitative codes across
the parallel analyses.

**Coding agreement**

The second author was assigned 30 randomly chosen articles containing 58
findings (18.6%), and coded them on all codes except authors and the specific CCFI
values. Mean agreement across the codes was 94.8% (range 82.8 [any dichotomous
indicators] to 100% [multiple codes]).

**Results**

The mean publication year was 2005.0 (range 1979-2011, but the rapid growth in
taxometric research illustrated in Figure 1 yields a median of 2007. Sample sizes for the
311 findings varied widely ($M=1,999.7$, median=934, range 130–80,304) and were log-transformed for later analyses. A summary of categorical finding-level codes is presented in Table 2, which indicates that most studies employed clinical, forensic, or undergraduate samples, used self-report data, and conducted MAMBAC and MAXCOV analyses with indicators of demonstrated validity. Methodological choices known to weaken indicator validity (i.e., single-item indicators) or bias findings towards taxonic conclusions (i.e., dichotomous indicators) were common, but taxometric procedures other than the MAMBAC, MAXCOV, MAXEIG, and L-Mode were rare. On average, 2.12 distinct taxometric procedures were employed in the analysis of each finding. CCFI values were reported or derivable for a substantial minority of findings.

A composite index of methodological quality was constructed from several codes in a way that subtracted points for known methodological weaknesses and added them for known strengths. One point was subtracted for findings based on samples of <300, using 1-item indicators, or using dichotomous indicators. One point was added for findings with an above-median sample size, an above-median number of distinct taxometric procedures, and demonstrated indicator validity. The five items composing this scale yielded an index ranging from -3 to +3 ($M=0.76$, $SD=1.56$), and all items intercorrelated positively ($\alpha = 0.67$).

Table 3 summarizes the distribution of taxonic findings and the mean CCFI values for the subset of findings reporting them. CCFI values derived from different procedures were highly similar (MAMBAC=0.37; MAXCOV=0.35; MAXEIG=0.36; L-Mode=0.36), and all intercorrelated strongly (mean $r=0.64$). A large minority (38.9%) of the 311 findings were taxonic and the mean CCFI for the 136 relevant findings was 0.37,
indicating that these findings tend to be predominantly nontaxonic. Indeed, the
distribution of mean CCFI values (see Figure 2), which reveals an apparent bimodality,
shows that only 20 (14.7%) findings exceeded the taxonic threshold (CCFI>0.5).
According to more conservative guidelines, 102 (75%) findings were clearly dimensional
(<0.4), 17 (12.5%) were clearly taxonic (>0.6), and 17 (12.5%) were ambiguous.

Table 3 indicates that the rate of taxonic findings differed widely across the
construct domains, $\chi^2_{(9)} = 56.73$, $p<.0001$. Taxonic findings were relatively infrequent in
the anxiety disorder, externalizing, other PD, and normal personality domains, and much
more common in the eating disorder, substance use, schizotypy, and miscellaneous
domains. Taxonic findings were also significantly more common in the broad
psychopathology arena (i.e., the combination of the mood, anxiety, eating, substance use,
externalizing, schizotypy, and other PD domains) than in the normal personality domain
(38.7% versus 16.3%; $\chi^2_{(1)} = 8.72$, $p<.01$.

Rates of use of the CCFI were also highly variable across construct domains,
$\chi^2_{(9)} = 68.90$, $p<.0001$, and the three domains with the highest rate of taxonic findings used
the CCFI the least. Evidence of a further disjunction between rates of taxonic findings
and CCFI evidence is apparent in Table 3. Only one of the four domains with a majority
of taxonic findings has a mean CCFI consistent with predominant taxonicity.

In a first attempt to assess factors related to taxonic findings, we examined
associations between taxonicity and three potential predictors. The findings reported
above suggest that use of the CCFI is negatively associated with taxonic findings. The
relative recency of the CCFI suggests that publication year might show the same negative
association. Finally, the methodological quality of studies, assessed independently of the
use of the CCFI, with which it correlated positively ($r=0.63$), might be associated with taxometric findings. Indeed, findings of analyses using the CCFI were much less likely to be taxonic (16.2%: 22 of 136) than those that did not (56.6%: 99 of 175), $\chi^2(1) = 52.54, p < 0.0001$. More recently published findings were also less likely to be taxonic ($r = -0.36, p < 0.001$) – 54.5% of findings published prior to the median article year were taxonic, compared to 23.2% of later findings– as were the findings of studies of higher methodological quality ($r = -0.33, p < 0.001$). In short, more recent, methodologically stronger research that systematically compares the fit of taxometric output to simulated taxonic and nontaxonic datasets is substantially less likely to yield taxonic findings than older and methodologically weaker research that does not.

These findings indicate that methodological and temporal factors are systematically associated with taxometric findings. However, because these predictors are related (e.g., more recent findings were methodologically stronger, $r = 0.68$ and more likely to use the CCFI, $r = 0.65$), and other methodological and construct factors also predict findings, multivariate analyses are required to assess their unique effects. The hierarchical structure of the data, with findings nested in articles, necessitates an analysis that takes these dependencies into account. We therefore employed multilevel logistic regression analysis (Bryk & Raudenbush, 1992; Snijders & Bosker, 1999).

**Multilevel logistic regression analyses**

We first ran single-predictor 2-level models with findings nested in articles. For ease of interpretation and because we are not primarily interested in within-article relationships between variables (Enders & Tofighi, 2007), binary predictors were entered
uncentered, and non-binary predictors as well as the level-2 predictor (publication year) were entered grand-mean centered. All analyses relied on population-average models, with random intercepts but no random slopes. The fixed effects are presented in Table 4. At level 2 (article), publication year was negatively associated with taxonicity: more recent studies were less likely to yield taxonic findings. At level 1 (finding), a number of methodological predictors were significantly associated with taxonic findings: findings based on smaller samples, fewer taxometric procedures, and indicators that were dichotomous, lacking reported validity, and based on single-items were more likely to be taxonic. These predictors all compose the methodological quality index, which was consequently strongly associated with nontaxonic findings. Findings in which the CCFI was reported were much less likely to yield taxonic findings, as were those based in part on the MAMBAC, MAXEIG, and L-Mode procedures. The findings based in part on the MAXCOV procedure and on “other” taxometric procedures were significantly more likely to be taxonic. Findings in the schizotypy and miscellaneous construct domains were more likely to be taxonic, and those in the externalizing domain were less likely.

Many successful predictors from these single-predictor analyses were correlated. For this reason, further analyses were conducted to tease apart unique effects. With a view to reducing the number of predictors in the final analysis the quality index was employed in place of the five predictors that constituted it. When the effects of the individual taxometric procedures were tested in analyses that also included quality, year, and use of the CCFI, only L-Mode (B=-1.144, SE=0.526, t(303)=2.18, p<.05) and MAXSLOPE (B=2.101, SE=0.875, t(303)=2.40, p<.05) were significant, so only these two procedure variables were retained for the final analysis. However the ten construct
domains were retained (represented as nine dummy variables with the “miscellaneous”
domain as the baseline), given the theoretical interest in determining which substantive
domains were more likely to contain taxa.

The final analysis therefore predicted taxonic effects with publication year at level
2 and quality, CCFI, L-Mode, MAXSLOPE, and nine construct domain dummies at level
1. The results (see Table 5) indicate that the methodological quality effects obtained in
the single-predictor analyses disappear when other variables are statistically controlled,
and that only use of the CCFI is significantly associated with a lower likelihood of
taxonic findings. Findings based at least in part on L-Mode were significantly less likely
to be taxonic and those based in part on MAXSLOPE were more likely. The construct
domain effects, which represent the likelihood of taxonic findings in each domain relative
to the miscellaneous domain, indicate that taxonic findings are less prevalent in the
domains of normal personality, abnormal personality (other than schizotypy), eating
disorders, and anxiety disorders, with marginal negative effects for mood disorders and
other individual differences. An analogous model with all variables grand-mean centered
(such that each fixed parameter reflects the effect relative to the average study), leads to
the same conclusions. Three of these effects were moderated by publication year: taxonic
findings were less likely to be obtained in more recent studies of normal personality (B=-
0.204, SE=0.103, \( t(283)=1.98, \ p<.05 \)), externalizing disorders (B=-0.455, SE=0.183,
\( t(283)=2.49, \ p<.05 \)), and schizotypy (B=-0.351, SE=0.104, \( t(283)=3.39, \ p<.005 \)).

Monte Carlo evidence (Ruscio & Kaczetow, 2009; Ruscio et al., 2010) indicates
that the CCFI is a highly accurate method for discriminating between taxonic and
dimensional data even under unfavourable measurement conditions. Our finding that use
of the CCFI is associated with a markedly reduced rate of taxonic findings implies that some taxonic findings in which the index was not used may be spurious. The same may be true of studies of lower methodological quality, which were more common in early taxometric research and more likely to deliver taxonic findings. To estimate the “true” prevalence of taxonicity in our sample of findings after controlling for these two potential sources of error, we ran a multilevel model using only the CCFI and quality index as predictors. The predictive equation, with a significant effect of CCFI ($B=-1.576, p<.001$) and a marginal effect of quality ($B=-0.201, p=.087$), was then used to estimate the probability of taxonic findings if the fit index was present (1) and quality was maximum (+3). The estimated probability (0.14) was markedly less than the proportion of taxonic findings reported in the literature (0.39). By implication, the taxometric literature substantially overstates the true frequency of taxa in its field of study.

**Discussion**

Our analysis of the taxometric literature has implications for taxometric methodology and for the latent structure of psychological phenomena. With regard to methodology, the analysis suggests that several factors that might plausibly influence taxometric research findings had no detectable effects. Different sample types – undergraduate, clinical, forensic, or child and adolescent – did not differ in rates of taxonic findings. The same was true of different data types – self-ratings, other-ratings, and interview-based judgments – despite some concerns that self-report data may produce pseudotaxonic findings (Beauchaine & Waters, 2003).
Other methodological factors that have been identified as potential sources of bias had systematic effects. Research using relatively small samples and few taxometric procedures was more likely to yield taxonic findings, consistent with Meehl’s (1995) view that modest samples are problematic and consistency testing essential. Use of dichotomous indicators was also associated with a higher rate of taxonic findings, as Ruscio (2000) contended, as was the use of indicator variables of questionable validity, such as those based on single items and those for which quantitative evidence of validity was not reported. Together these factors formed a reliable index of methodological quality which powerfully predicted rates of taxonic findings: methodologically stronger research tended not to find taxa. Unexpectedly two taxometric procedures were also associated with taxonic findings, L-Mode negatively and MAXSLOPE positively. The latter effect may have limited importance because MAXSLOPE is rarely used.

Our review points to historical changes in taxometric practice. Methodological quality has improved over time, with recent findings more likely to use several distinct taxometric procedures, less likely to use small samples, and less likely to employ indicators that are dichotomous, single-item, or lacking demonstrated validity. This rise in methodological quality has accompanied, and probably partly driven, a decline in taxonic findings. Clearly the most important historical development in taxometric practice has been the analysis of simulated comparison data and use of the CCFI (Ruscio et al., 2007). This technique has rapidly become dominant in taxometric research, serving as an objective supplement or alternative to visual curve inspection, and our analysis found it to be a powerful predictor of structural findings. Research that did not use the technique was almost 3.5 times as likely to yield taxonic findings as research that did
(56.6% versus 16.2%), and this effect persisted even when potentially confounding methodological and construct-related factors were statistically controlled.

The use of simulated comparison data has been controversial, and it might be argued that its negative association with taxonic findings reflects a pro-dimensional bias rather than enhanced validity. We believe this argument is baseless. First, use of the CCFI was strongly and positively associated with methodological quality, implying that it is a positive adjunct to research that is methodologically stronger in other respects. Second, research that was methodologically stronger in uncontroversial ways (e.g., larger samples, use of more convergent data-analytic procedures) was also associated with lower rates of taxonic findings. Third, substantial Monte Carlo evidence (Ruscio et al., 2007; Ruscio & Kaczetow, 2009; Ruscio et al., 2010) attests to the CCFI’s high accuracy, its robustness over a wide range of unfavourable measurement conditions that occur in real research settings, and its lack of significant bias either for or against taxonic findings. For example, Ruscio et al.’s (2010) analysis of 100,000 simulated datasets found that the CCFI correctly identified taxonic and dimensional data with equally high accuracy.

The re-orientation of taxometric research in favour of dimensional models that the simulated comparison data technique has brought about is substantial. Our findings indicate that methodologically stronger research that uses the CCFI is particularly unlikely to find taxa, whereas early, pre-CCFI taxometric research, which appeared to find many taxa, tended to be methodologically weaker. By implication, early taxometric research is likely to contain invalid taxonic findings. Our analysis estimated the rate of taxonic findings in our 177 articles to be 14% rather than the observed 39%, had the CCFI always been used in studies of high methodological quality. If this estimate is
accurate, almost two thirds of reported taxonic findings are invalid. Moreover, most of the invalid findings would be found among the 56.3% of findings (175 of 311) in which the CCFI was not employed. The reported rate of taxonic findings in these studies was 56.6% (99 of 175), so if the “true” rate of taxonicity was approximately 14% then three out of four taxonic findings generated by non-CCFI studies were probably spurious.

The way in which research using simulated comparison data and the CCFI has challenged earlier taxonic findings is well illustrated by anxiety sensitivity and antisociality (i.e., antisocial conduct, antisocial PD, criminal lifestyle and thinking styles, and psychopathy). The first seven articles on anxiety sensitivity, published between 2005 and 2007, did not use the CCFI and generated 11 exclusively taxonic findings. The next three, published 2008 to 2011, used the CCFI and generated 7 exclusively nontaxonic findings (mean CCFI=0.34). An equally stark reversal occurred for antisociality: four early articles (1994-2005) that did not use the CCFI yielded exclusively taxonic findings, whereas 17 later articles that did (2004-2011) yielded 18 consistently nontaxonic findings (all CCFIs<.04, mean=0.28), The CCFI data-analytic strategy can therefore sharply and replicably challenge early taxonic findings.

If the “true” rate of taxonic findings in the 177 taxometric articles is only about 14% then taxonic latent structure is rather rare in psychology and psychiatry, instead of being common as the 39% figure would suggest. If it is further assumed that taxometric researchers have selectively examined latent variables for which taxonic structure was plausible, and have attempted to replicate findings where such structure was initially supported (e.g., the 29 findings on schizotypy), then taxa may be even rarer in the domains of personality and psychopathology than the former figure suggests.
This conclusion regarding dimensional versus categorical latent variation complements the findings of a recent meta-analysis of continuous versus discrete observed variation (Markon, Chmielewski & Miller, in press). Continuous measures of psychopathology were on average 15% more reliable and 37% more valid than discrete measures, a superiority that was invariant across different construct types. The authors proposed that “in the absence of a specific rationale for the contrary, continuous measures of psychopathology should be preferred over discrete measures a priori.” Our findings suggest that dimensional models of the latent structure of personality and psychopathology should similarly be preferred over categorical models a priori, with the proviso that if 14% of taxometric findings are validly taxonic then there may be specific rationales to the contrary in some cases. Our findings certainly do not justify blanket claims that all psychological variation is latently continuous.

The findings of this review help to identify where latent discontinuities exist. They strongly suggest, as expected, that taxa are unlikely to be found in the domains of normal personality and PDs, supporting the standard assumptions of trait theory on the one hand and much recent PD scholarship on the other (e.g., Clark, 2007). The findings also suggest that taxa are scarce within the internalizing spectrum, such as in the mood, anxiety, and eating disorders, and also in the general externalizing domain. The only domains in which taxonic findings were at least somewhat prevalent were schizotypy and substance use disorders. Neither set of findings was consistently taxonic, the substance use domain was not significantly associated with taxonicity in the single-predictor analysis, and neither domain yielded taxonic findings at a significantly higher rate compared to the miscellaneous domain or to the average study in the final analysis.
Nevertheless, these two domains stood out as offering at least some support for the existence of taxa. The schizotypy domain finding is perhaps unsurprising given its long record of taxonic findings, dating back to Golden and Meehl (1979). Both findings qualify the conclusion that abnormal personality and externalizing disorders tend to be latently continuous: schizotypy and substance use may be at least partial exceptions to these generalizations.

In deciding which specific latent variables may be taxonic, it seems prudent to give special weight to those taxonic findings in which the CCFI was employed, given the apparent taxonic bias of research in which it was not. Fifteen such findings (excluding two from the miscellaneous domain) exceed a conservative criterion for taxonicity (CCFI>0.6). These findings predominantly represent the domains of substance use (7) and other individual differences (4). The domains of mood (depression; Ruscio et al., 2009), anxiety (social anxiety disorder; Weeks et al., 2010), eating (binge-eating disorder; Hilbert et al., 2011), and externalizing disorders (intermittent explosive disorder; Ahmed et al., 2010) are represented by one finding each. These solo findings do not provide a strong basis for inferring taxa as they are either unreplicated or run contrary to many nontaxonic findings for the same construct.

The existence of substance use taxa is supported by elevated CCFI values in five studies of alcohol abuse, misuse, or dependence (Green et al., 2011; Walters, 2008; Walters, 2009; Walters, Hennig et al., 2009; Walters & Ruscio, 2010) and in two studies of nicotine dependence (Goedeker & Tiffany, 2008). Although several studies using the CCFI have yielded nontaxonic findings in this domain (alcohol abuse and dependence: Slade, Grove, & Teeson, 2009; Walters, Diamond, & Magaletta, 2010; cannabis
there is at least some replicated evidence that substance use taxa exist.

In the “other individual differences” domain, three of the four strongly taxonic findings relate to autism and its subtypes (Frazier et al., 2010; Ingram et al., 2008). Although two of these findings came in the context of a study in which five other possible subtypes appeared to be nontaxonic (Ingram et al., 2008), there now appears to be consistent support for the existence of taxa in the autism realm, including an additional study that did not use the CCFI (Munson et al., 2008).

In sum, after 32 years of taxometric research, the methodologically strongest evidence for the existence of taxa is in the areas of alcohol- and nicotine-related substance use and autism. Schizotypy should probably be added because a clear majority of the findings addressing it have been taxonic, although the only ones to employ the CCFI have not (Rawlings, Williams, Haslam, & Claridge, 2008). We therefore suggest that the taxometric literature clearly supports the existence of taxa only in the areas of schizotypy, autism, and substance use disorders. Ideally, researchers should go beyond identifying such taxa and demonstrate that they have construct validity, such as predicted patterns of association with external variables (Hasin et al., 2006; Prisciandaro & Roberts, 2011), in the absence of which evidence for taxonicity can be questioned (Watson, 2003). We also acknowledge that taxometrics is just one statistical approach to the question of latent structure, and any comprehensive analysis of the existence and distribution of latent categories in personality and psychopathology must integrate its findings with those of other approaches, which may systematically differ.
Conclusions

Taxometrics was launched as a rigorous method for “solving” (Meehl, 1995) a fundamental question in the study of personality and psychopathology. Three decades of research allow us to review its answers. The collected evidence of this review is that latent categories are very scarce in normal and abnormal personality and infrequent but not absent in the broad arena of psychopathology. Taxa are most likely few and far between in the domains of internalizing and externalizing disorders, with the partial exception of substance use conditions. They may also occur in additional forms of psychopathology, notably schizotypy and autism. Other taxa may exist, but taxometric evidence for them is preliminary and runs against the strongly dimensional tide of recent research. Early taxometric research findings performed a valuable role in raising the issue of latent structure and giving credence to the possibility of psychological taxa, but their findings may often have been spurious. Taxometric research supports the conclusion that most psychological variation is dimensional, with a few potentially important exceptions.
References (* used in review)


Developmental Psychology 39, 387-404.


## Classification of constructs examined in taxometric research.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Constructs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mood disorder</td>
<td>Depression &amp; its subtypes, Depression-proneness, Dysthymia, Grief, Mania, Mixed anxiety-depression</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td>Agoraphobia, Anxiety, Anxiety sensitivity, Aversion, Disgust sensitivity, Distress, Fear of pain, Health anxiety, Hypochondriasis, OCD &amp; subtypes, PTSD, Separation anxiety, Worry (normal &amp; pathological)</td>
</tr>
<tr>
<td>Eating disorder</td>
<td>Anorexia nervosa &amp; subtypes, Binge eating disorder, Body dissatisfaction, Bulimia nervosa &amp; subtypes, Dietary restraint, Drive for thinness, Eating pathology</td>
</tr>
<tr>
<td>Substance use</td>
<td>Alcohol abuse &amp; dependence, Cannabis dependence, Nicotine addiction/dependence</td>
</tr>
<tr>
<td>Externalizing</td>
<td>Adolescent externalizing, Aggression &amp; subtypes, Antisocial behavior, Antisocial PD, Criminal lifestyle, Criminal thinking style, Intermittent explosive disorder, Internet gambling, Psychopathic sexuality, Psychopathy, Risky sexual behavior, Sexual violence risk</td>
</tr>
<tr>
<td>Schizotypy</td>
<td>Pre-schizophrenic personality, Schizoidia, Schizophrenia risk, Schizotypy &amp; subtypes</td>
</tr>
<tr>
<td>Other personality</td>
<td>Avoidant PD, Borderline PD, Dependent PD, Depressive PD, Narcissistic PD, Obsessive-compulsive PD, Paranoid PD</td>
</tr>
<tr>
<td>Normal personality</td>
<td>Alexithymia, Attachment styles, Child temperament dimensions, Extraversion/introversion, Femininity, Hypnotic susceptibility, Hypomaniac temperament, Impulsivity, Infant reactivity, Jungian temperament dimensions, Narcissism, Perfectionism, Self-monitoring, Sexual orientation, Type A, Type D</td>
</tr>
<tr>
<td>Category</td>
<td>Conditions</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Other individual difference</td>
<td>ADHD &amp; subtypes, Adjudicative competence, Autism &amp; subtypes, Dementia, Cognitive symptom exaggeration, Dissociation, Feigned neurocognitive deficit, Health complaint exaggeration, Hypersexuality, Impression management, Infrequency responding, Language impairment, Malingering, Psychosis &amp; subtypes, Racism, Schizophrenia &amp; subtypes, Self-deceptive positivity, Somatic complaints, Somatization, Symptom over-reporting</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>Biological sex, Envy/jealousy, Handedness, Marital discord, Metabolic syndrome, Nociceptive flexion reflex, Prostate cancer risk, Relationship types, Tardive dyskinesia</td>
</tr>
</tbody>
</table>
Table 2.

Descriptive statistics on finding-level codes (N=311)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical/forensic sample</td>
<td>125 (40.2)</td>
</tr>
<tr>
<td>Undergraduate sample</td>
<td>115 (37.0)</td>
</tr>
<tr>
<td>Community sample</td>
<td>86 (27.7)</td>
</tr>
<tr>
<td>Child/adolescent sample</td>
<td>54 (17.4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data type</th>
<th>Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-ratings</td>
<td>212 (68.2)</td>
</tr>
<tr>
<td>Interview</td>
<td>67 (21.5)</td>
</tr>
<tr>
<td>Other-ratings</td>
<td>23 (7.4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Indicator construction</th>
<th>Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicator validity demonstrated</td>
<td>199 (64.0)</td>
</tr>
<tr>
<td>Single-item indicators</td>
<td>126 (40.5)</td>
</tr>
<tr>
<td>Dichotomous indicators</td>
<td>62 (19.9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data analysis</th>
<th>Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAMBAC</td>
<td>234 (75.2)</td>
</tr>
<tr>
<td>MAXCOV</td>
<td>193 (62.1)</td>
</tr>
<tr>
<td>MAXEIG</td>
<td>124 (39.9)</td>
</tr>
<tr>
<td>L-Mode</td>
<td>93 (29.9)</td>
</tr>
<tr>
<td>MAXSLOPE</td>
<td>5 (1.6)</td>
</tr>
<tr>
<td>Other taxometric procedure</td>
<td>12 (3.9)</td>
</tr>
<tr>
<td>CCFI employed</td>
<td>136 (43.7)</td>
</tr>
</tbody>
</table>
Table 3.
Distribution of taxonic findings and use of the CCFI across construct types

<table>
<thead>
<tr>
<th>Grouping</th>
<th>Findings</th>
<th>Taxonic (%)</th>
<th>Number (%)</th>
<th>CCFI</th>
<th>Mean CCFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mood disorder</td>
<td>36</td>
<td>13 (36.1)</td>
<td>12 (33.3)</td>
<td>.39</td>
<td></td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td>60</td>
<td>16 (26.7)</td>
<td>32 (53.3)</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td>Eating disorder</td>
<td>21</td>
<td>13 (61.9)</td>
<td>4 (19.0)</td>
<td>.40</td>
<td></td>
</tr>
<tr>
<td>Substance use</td>
<td>12</td>
<td>7 (58.3)</td>
<td>12 (100)</td>
<td>.57</td>
<td></td>
</tr>
<tr>
<td>Externalizing</td>
<td>29</td>
<td>6 (20.3)</td>
<td>24 (82.8)</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>Schizotypy</td>
<td>29</td>
<td>21 (72.4)</td>
<td>4 (13.8)</td>
<td>.37</td>
<td></td>
</tr>
<tr>
<td>Other PD</td>
<td>12</td>
<td>1 (8.3)</td>
<td>9 (75.0)</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>Normal personality</td>
<td>41</td>
<td>8 (19.5)</td>
<td>11 (26.8)</td>
<td>.34</td>
<td></td>
</tr>
<tr>
<td>Other individual difference</td>
<td>49</td>
<td>18 (36.7)</td>
<td>24 (49.0)</td>
<td>.41</td>
<td></td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>22</td>
<td>18 (81.8)</td>
<td>4 (18.2)</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>311</td>
<td>121 (38.9)</td>
<td>136 (43.7)</td>
<td>.37</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.

Findings of single-predictor multilevel logistic regression analyses predicting taxonic findings

<table>
<thead>
<tr>
<th>Predictor</th>
<th>B (SE)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 2 (df=175)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Publication year</td>
<td>-0.133 (0.028)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Level 1 (df = 309)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N (log-transformed)</td>
<td>-0.331 (0.143)</td>
<td>.022</td>
</tr>
<tr>
<td>Undergraduate sample</td>
<td>0.053 (0.295)</td>
<td>.858</td>
</tr>
<tr>
<td>Clinical/forensic sample</td>
<td>0.110 (0.271)</td>
<td>.683</td>
</tr>
<tr>
<td>Community sample</td>
<td>0.148 (0.273)</td>
<td>.589</td>
</tr>
<tr>
<td>Child/adolescent sample</td>
<td>0.176 (0.353)</td>
<td>.618</td>
</tr>
<tr>
<td>Data type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-ratings data</td>
<td>0.282 (0.289)</td>
<td>.923</td>
</tr>
<tr>
<td>Other-ratings data</td>
<td>-0.472 (0.504)</td>
<td>.350</td>
</tr>
<tr>
<td>Interview data</td>
<td>-0.242 (0.316)</td>
<td>.444</td>
</tr>
<tr>
<td>Number of indicators</td>
<td>0.089 (0.056)</td>
<td>.110</td>
</tr>
<tr>
<td>1-item indicators</td>
<td>0.669 (0.286)</td>
<td>.020</td>
</tr>
<tr>
<td>Dichotomous indicators</td>
<td>1.061 (0.387)</td>
<td>.007</td>
</tr>
<tr>
<td>Validated indicators</td>
<td>-0.600 (0.299)</td>
<td>.045</td>
</tr>
</tbody>
</table>

Data analysis
<table>
<thead>
<tr>
<th>Metric</th>
<th>Value (SE)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAXCOV</td>
<td>0.637 (0.290)</td>
<td>.029</td>
</tr>
<tr>
<td>MAMBAC</td>
<td>-1.601 (0.322)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>MAXEIG</td>
<td>-0.712 (0.291)</td>
<td>.015</td>
</tr>
<tr>
<td>L-Mode</td>
<td>-1.870 (0.335)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>MAXSLOPE</td>
<td>1.755 (1.069)</td>
<td>.101</td>
</tr>
<tr>
<td>Other</td>
<td>2.372 (0.835)</td>
<td>.005</td>
</tr>
<tr>
<td>Number of procedures</td>
<td>-0.934 (0.189)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Comparative fit index</td>
<td>-1.920 (0.303)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Overall study quality</strong></td>
<td>-0.468 (0.104)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

**Construct domain**

<table>
<thead>
<tr>
<th>Category</th>
<th>Value (SE)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mood disorder</td>
<td>-0.126 (0.387)</td>
<td>.745</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td>-0.670 (0.397)</td>
<td>.092</td>
</tr>
<tr>
<td>Eating disorder</td>
<td>0.747 (0.612)</td>
<td>.223</td>
</tr>
<tr>
<td>Substance abuse</td>
<td>0.842 (0.671)</td>
<td>.211</td>
</tr>
<tr>
<td>Externalizing</td>
<td>-0.954 (0.450)</td>
<td>.035</td>
</tr>
<tr>
<td>Schizotypy</td>
<td>1.729 (0.517)</td>
<td>.001</td>
</tr>
<tr>
<td>Other PD</td>
<td>-1.714 (0.940)</td>
<td>.069</td>
</tr>
<tr>
<td>Normal personality</td>
<td>-0.949 (0.582)</td>
<td>.104</td>
</tr>
<tr>
<td>Other individual difference</td>
<td>-0.019 (0.378)</td>
<td>.959</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>1.855 (0.683)</td>
<td>.007</td>
</tr>
</tbody>
</table>
Table 5.

Findings of final multilevel logistic regression analysis predicting taxonic findings

<table>
<thead>
<tr>
<th>Predictor</th>
<th>B</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication year</td>
<td>0.039 (0.062)</td>
<td>.532</td>
</tr>
<tr>
<td>Comparative fit index</td>
<td>-3.446 (0.881)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Methodological quality</td>
<td>0.132 (0.173)</td>
<td>.447</td>
</tr>
<tr>
<td>L-Mode</td>
<td>-1.391 (0.538)</td>
<td>.011</td>
</tr>
<tr>
<td>MAXSLOPE</td>
<td>2.561 (0.741)</td>
<td>.001</td>
</tr>
<tr>
<td>Construct domain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mood disorder</td>
<td>-1.764 (0.903)</td>
<td>.051</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td>-2.168 (0.884)</td>
<td>.015</td>
</tr>
<tr>
<td>Eating disorder</td>
<td>-1.867 (0.905)</td>
<td>.040</td>
</tr>
<tr>
<td>Substance abuse</td>
<td>0.259 (2.341)</td>
<td>.912</td>
</tr>
<tr>
<td>Externalizing</td>
<td>-0.510 (0.893)</td>
<td>.568</td>
</tr>
<tr>
<td>Schizotypy</td>
<td>-1.189 (0.894)</td>
<td>.185</td>
</tr>
<tr>
<td>Other PD</td>
<td>-2.966 (1.116)</td>
<td>.009</td>
</tr>
<tr>
<td>Normal personality</td>
<td>-4.162 (1.030)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Other individual difference</td>
<td>-1.684 (0.866)</td>
<td>.052</td>
</tr>
</tbody>
</table>
Figure 1: Publication year of the 177 articles
Figure 2. Distribution of mean CCFI values for findings reporting them ($n=136$)