Hunting for Snarks: A Reply to McPhail (2019)

Author Note

Abstract

No Abstract is required.

Keywords: Taxometrics, Sexual Sadism, Latent Class, Agonistic Continuum, Paraphilic Coercive Disorder.

In 2018, we examined the latent structure of sexual sadism and published our results in *Archives of Sexual Behavior*. Because sadism is treated as a disorder that one does or does not have, sadists have been conceptualized as fundamentally different from non-sadists, and research is directed at a very specific subgroup of offenders (Longpré, Guay, Knight, & Benbouriche, 2018). Sadism has, however, been plagued with an absence of pathognomonic symptoms, and many symptoms supposedly identifying sadism have been found among non-sadistic offenders and non-delinquents (for more details, see Longpré et al., 2018). Based on the results obtained in Longpré et al. (2018), but also on the convergence of results across studies and samples, we concluded that sexual sadism is distributed as a dimension. These results are not only in stark contrast to the categorical orientation that has been and continues to be supported in the Diagnostic and Statistical Manual (e.g., American Psychiatric Association, 2013), but they also provide empirical evidence for the importance of reformulating existing categorical diagnoses of sadism. If a construct is distributed dimensionally, cutoffs are arbitrary and have to be validated empirically to maximize the objectives of particular assessments. For such constructs, no naturally occurring demarcation exists that can identify a distinct group that differs in kind and can justify clinical or forensic decisions.

In a Letter to the Editor, McPhail (2019a) challenged our conclusions and proposed that a trichotomous "ghost" haunted our data. McPhail (2019a) argued that L-Mode curves revealed a trichotomous structure, similar to his findings on pedophilia (McPhail, Olver, Brouillette-Alarie, & Looman, 2018), and that our results should be reinterpreted accordingly. Although some interesting distinctions among those who sexually abuse children emerged in McPhail et al.'s (2018) pursuit of a trichotomous categorization of pedophila, research on sexual sadism and on taxometrics do not support McPhail's (2019a) interpretation. The core of the disagreement

between McPhail's (2019a) and our strategies for interpreting the results of taxometrics focuses on the degree to which one gives interpretive credence to selected graphic representations of particular analyses rather than, as we argue, to the overall consistency of results across methods and studies. We also believe, in contrast to McPhail (2019a), that the results of Monte Carlo studies, which provide helpful parameters to assist optimal decisions about taxometrics, should guide interpretations. McPhail's (2019a) reinterpretation of our results was both flawed and erroneous, involving: (1) an implicit rejection of Meehl's basic theoretical strategy that looks for numerical consistency over different procedures, (2) an insufficient attention both to the guidelines provided by studies that explore the methodological limitations of taxometrics and to the convergence of results across empirical studies, and (3) a failure to take into account recent research on the latent structure of the Agonistic Continuum. We address each of these errors in turn.

Rejecting the Fundamental Features of Meehl's Taxometrics

Meehl's solution of the classification problem is an instantiation of both his Neo-Popperian philosophy of science and his belief that construct validity is essential to the advancement of science (Cronbach & Meehl, 1956; Meehl, 1990, 1995). The former demands that we put hypotheses at severe risk for disconfirmation, and the latter requires consistent converging lines of evidence across multiple variables, multiple independent ways of measuring a construct, and multiple different samples or populations. The semi-independent mathematical definitions of taxonicity that Meehl (1995) created for his taxometrics constitute severe tests of hypotheses about the distribution of constructs. Coherence among nonredundant indicators provides strong corroboration of either taxonic or dimensional distribution. In his solution, consistency tests are essential. Their absence would lead to excessive reliance on fallible

judgment and to a lack of consensus among scientists. The practice that McPhail (2019a, b) proposes of allowing the interpretation of particular selected taxometric figures to trump the strategy of assessing numerical consistency across several semi-independent algorithms is fraught with problems and is a rejection of the fundamental core of Meehl's solution.

Initially, Meehl hypothesized that the graphical representation of how a construct performed in each of his taxonomic algorithms would be sufficiently unambiguous that unbiased observation of their shapes would be sufficient to determine the outcome direction of each taxon or dimension (e.g., Meehl & Yonce, 1994). It turned out, however, that problems of the skewness of a construct's distribution, the vagaries of psychometric properties of the indicators used, and the baserate of the purported taxon substantially affected the shape of distributions that resulted. Even though Meehl clearly recognized the problem of fallible judgment, he underestimated the degree to which researchers' confirmatory bias and "p-hacking" tendencies would affect their judgments. Ruscio, Haslam, and Ruscio (2006) and Ruscio, Ruscio, and Meron (2007) rectified this interpretive problem by developing a procedure in which they constructed optimal taxonic and dimensional comparison curves using a bootstrapping technique that accounted for the unique distributional and correlational characteristics of a particular data set. They quantified the relative fit between the actual data and these two simulated comparison curves with the comparison curve fit index (CCFI). Monte Carlo studies support the CCFI's accuracy (Ruscio, 2007; Ruscio & Marcus, 2007; Ruscio et al., 2007; Ruscio, Walters, Markus, & Kaczetow, 2010; Walters, McGrath, & Knight, 2010; Walters & Ruscio, 2009, 2010). Interpretation of the graphs should never take precedence over the CCFIs, which embody the fundamental principles upon which Meehl's taxometrics were based and have repeatedly been shown to be accurate. Doing so leads to the errors in McPhail's (2019a) work (see also McPhail et al., 2018).

Specific Problems in McPhail's (2019) Interpretation

Insufficient Attention to Methodological Studies of Taxometrics

McPhail (2019a) criticized Longpré et al.'s indicators, suggesting that indicator skew and the low baserate of the purported taxon may have masked an unseen taxon. Ruscio and Marcus (2007) found data that supported the use of taxometrics with skewed indicators and baserates as low as .05. The empirically derived baserate in Longpré et al. was around .20, considerably above the baserate that should elicit concern. Some nuisance covariation and low discrimination between the putative taxon and complement groups in Longpré et al.'s indicators may have contributed to suppressing the discrimination between simulated comparison curves, but examination of all figures indicates moderately good differential discrimination. Nonetheless, G. D. Walters (personal communication, September 3, 2019) suggested a strategy of dropping the two weakest indicators and rerunning the taxometrics to address these minor problems. These reanalyses resulted in smoother curves and better discrimination between optimal taxonic and dimensional solutions, but there was little change in CCFIs. The apparent trimodal L-Mode, however, remained.

McPhail (2019a) placed heavy emphasis on two Monte Carlo studies of the taxometrics of polytomous constructs (McGrath, 2008; Walters et al., 2010). McPhail (2019a) neglected, however, to apply the overarching findings of these studies to Longpré et al.'s (2018) data. Because McGrath's (2008) exploration of polytomous outcomes was compromised by relying on a restricted set of sample parameters and by using taxonomic-dimensional indicators that have performed poorly in Monte Carlo research (e.g., Haslam & Cleland, 2002; Ruscio, 2007), Walters et al. (2010) attempted to address these problems using Ruscio's CCFIs. Their results were very clear. For the three-class samples, the mean CCFI was >.55 in 100% of the analyses

and >.60 in 98.9%. Thus, taxometrics almost always identified the discontinuity present in the three-class samples. Ruscio and Ruscio's (2004) two-step procedure, which attempted to corroborate a three-class distribution by exploring the potential for a two-taxon solution within an identified taxon, suffered from a significant specificity problem, with between 52% and 60% of the two-class samples incorrectly being identified as having a third type. Walters et al. (2010) concluded that their data had not provided a definitive answer to the question of whether taxometrics analysis can be used to differentiate between dichotomous and polytomous categorical structures. The principal strength of taxometrics was their ability "to identify discontinuity in a distribution accurately" (p. 153). Their principal weakness was their "inability to differentiate between two- and three-class samples" (p. 153). They argued that the problem of polytomy requires further research to develop appropriate, valid identification procedures. They did not advocate "ghost" hunting in L-Mode curves.

Using 50,000 Monte Carlo data sets, McGrath and Walters (2012) extended the prior Monte Carlo studies, assessing whether taxometrics can distinguish between dimensional (1-class) and categorical (2–5 classes) latent structures and exploring the efficacy of strategies attempting to estimate the number of classes in categorical datasets. Their results revealed that the CCFI did a significantly better job of identifying latent structure type than finite mixture modeling (a cluster analytic approach). Consistent with Walters et al. (2010), the CCFI correctly identified 100% of the 1-class data sets as continuous/dimensional and 99% of the 2- to 5-class data sets as categorical/taxonic. They proposed a second step after taxometrics to determine a latent structure model (either the number of dimensions or categories). Because, as in Walters et al. (2010), the CCFI was not particularly effective in identifying the number of categories in the polytomous data sets, McGrath and Walters (2012) recommended that researchers use an

alternative procedure like finite mixture modeling to determine the potential number of categories once a construct had been identified as taxonic during the first step.

Most telling in these Monte Carlo studies for McPhail's search for polytomy in Longpré et al.'s (2018) data is the finding that whereas three-class samples in Walters et al. (2010) and all polytomous samples in McGrath and Walters (2012) almost always gave rise to taxonic results across different taxometric procedures, there was no evidence at all of discontinuity in any procedures calculated in Longpré et al. (2018). The individual MAMBAC, MAXEIG, and L-Mode CCFIs were all below .45, and the mean CCFI was less than .30. Even in McGrath's (2008) study, MAXEIG consistently yielded taxonomic outcomes for the three-class samples. Consequently, the strong support across procedures for dimensionality in Longpré et al.'s (2018) results should preclude cherry-picking a single graphic representation in an attempt to support a post hoc speculation of a three-class solution. As both Walters et al. (2010) and McGrath and Walters (2012) concluded, Meehl's taxometrics do identify the presence of discontinuity, and thus the lack of any evidence for discontinuity in Longpré et al.'s results strongly supports the hypothesis that sexual sadism is distributed dimensionally.

Critiquing the trichotomous conclusion of McPhail et al. (2018) is beyond the scope of the current response. Suffice it to say that the L-Mode graph presented in support of trimodality was one of 12 L-Mode analyses done on different samples (see McPhail, 2019b). Although it was chosen as the best example of a trichotomous solution, even its purported "ghostly" trimodal appearance is clearly in the eye of the beholder. A sorting task that included many one-, two-, and three-class Monte Carlo generated curves would not likely find this curve in the three-class bin. Moreover, none of the remaining 11 L-Mode figures evidenced any trimodality. Also, in their post hoc analyses to explore the potential for a trichotomous solution for pedophilia,

McPhail et al. (2018) used Ruscio and Ruscio's (2004) follow-up taxometric procedure despite Walters et al.'s (2010) and McGrath and Walters' (2012) findings that such taxometrics are suboptimal for the Stage 2 identification of the number of categories after discontinuity has been found. McGrath and Walters (2012) recommended further that alternative procedures such as finite mixture models might be better suited to the task than taxometrics, but this is speculative and requires further research. They do not recommend searching L-Mode graphs for trimodal solutions.

Inadequate Consideration of Converging Evidence for the Dimensionality of Sadism

Consistency across samples and populations is also an important consideration when interpreting taxometrics. In response to McPhail's criticisms, we reexamined the taxometric curves in both the Knight, Sims-Knight, and Guay (2013) and Longpré, Sims-Knight, Neumann, Guay, and Knight (2019c) studies. No potential trimodal structures were evident, either on the full scale or on the sub-scales, with a baserate of 10% and 15%, for MAMBAC, MAXEIG, and L-Mode curves (all curves available upon request). In all the taxometric studies of sexual sadism (Knight et al., 2013; Longpré, Knight, & Guay, 2019b; Longpré et al., 2019c; Mokros, Schilling, Weiss, Nitschke, & Eher, 2014), only Longpré et al. (2018) produced an L-Mode that could potentially be interpreted as trimodal.

McPhail (2019a) also criticized the .45/.55 indeterminate threshold criterion used in Longpré et al. (2018) and thought that the interpretation of the CCFIs was too generous. McPhail (2019a) proposed that using a more "conservative" threshold of .40/.60 would produce ambiguous results that would go against the "liberal" conclusion of a dimension distribution. Ruscio et al. (2010) used Monte Carlo simulation to assess the accuracy of different thresholds. They found that using a "liberal" indeterminate criterion of .45/.55 yielded an accuracy rate of

98.2% for MAMBAC analysis, 95.8% for MAXEIG analysis, 97.3% for L-mode analysis, and 99.4% for mean CCFIs. In contrast, using a "conservative" indeterminate criterion of .40/.60 yielded accuracy rates of 99.2% for MABMAC, 97.9% for MAXEIG, 98.9% for L-Mode, and 99.85% for mean CCFIs. Although the increase in accuracy of the "conservative" over the "liberal" indeterminate criterion was trivial, the cost of the former relative to using no indeterminate criterion was an increase in unclassified cases of 14%, whereas the latter yielded only a 5% increase in unclassified cases, supporting the detection superiority of the latter. Ruscio et al. (2010) concluded:

> it is difficult to conceive of circumstances under which one would be willing to absorb such a substantially increased risk of ambiguous results in exchange for such a slight gain in accuracy (p. 19).

McPhail (2019a; McPhail et al., 2018) nonetheless prefers the more conservative criterion, and this choice impacts his interpretation of studies. For instance, McPhail (2019a) interprets both Knight et al.'s (2013) and Mokros et al.'s (2014) CCFI results as "falling between .400 and .600," and therefore being "ambiguous and not supportive of either taxonic or dimensional structure." Using instead the .45 to .55 threshold for the Knight et al. (2013) study indicates that only one (Killing) of the ten mean CCFIs (calculated for both 10% and 15% baserate estimates) exceeded .45 and here only barely (.458). Moreover, the mean CCFIs for the Agonistic Continuum, which captures the full range of sexual sadism, were .380 and .375, for the 10% and 15% baserate analyses, respectively. Even using McPhail's preferred conservative criteria, six of the ten mean CCFIs were in the dimensional direction, and none even approached the .50, .55, or .60 criteria for taxonicity. It is interesting that when the majority, but not all CCFIs were in a taxonic direction in their pedophilia study, McPhail et al. (2018) were quite

willing to argue for a taxonic outcome, but McPhail (2019a) was unwilling to accept this same decision standard when applying the conservative indeterminate criterion to Knight et al.'s (2013) results. Clearly, the overwhelming direction of Knight et al.'s results was dimensional, without any support for taxonicity. These results have been replicated in Longpré et al. (2019c). Moreover, a similar criticism about McPhail's interpretation of Mokros et al.'s (2014) results applies. Their mean CCFI (.447) falls in the dimensional range for the .45 to .55 thresholds. It is also noteworthy that in a sorting task for one-, two-, and three-class curves, there is little doubt that Mokros et al.'s L-Mode graph would likely be sorted into the one-class bin and would never be sorted into a three-class bin.

The Agonistic Continuum: Sexual Coercion as a Dimension

The research on the structure of sadism is consistent with recent studies on the Agonistic Continuum. The Agonistic Continuum is hypothesized to range from coercive fantasies, through fantasies of forcing sexual compliance, to fantasies and behaviors of hurting, humiliating, torturing, and killing. The Agonistic Continuum first emerged as a single dimension from factor analyses (exploratory and confirmatory), item response theory analyses, and taxometrics on a sample of sexual offenders (Knight et al., 2013). These results have subsequently been replicated and extended in new sample of sexual offenders (Longpré et al., 2019c), in a sample of college students, in a sample of college students and community individuals, and in a sample of college students, community individuals, violent offenders, and sexual offenders (Longpré et al., 2019b). Across these studies, no taxonic peaks have emerged at any of the Agonistic levels (PC+, Bondage, Beating, and Killing) either among these diverse samples (college students, community individuals, violent offenders, and sexual offenders) or between genders.

McPhail (2019a) cited Reale's (2017) cluster analytic results as support for his hypothesis of a three-class solution for sadism. The cluster analytic technique that Reale (2017) used, however, was suboptimal for identifying taxonic differences. As Walton, Ormel, and Krueger (2011) have indicated, classes that emerge from cluster analyses can represent degrees of severity of an underlying continuum as opposed to distinct taxonic groups. Although thresholds can always be found using cluster analyses, the resultant sub-groups are sometimes better conceptualized as differing along a continuum (Lahey & Waldman, 2003) rather than as having natural boundaries (Ruscio et al., 2006). It is noteworthy that contrary to McPhail's interpretation, Reale (2017; see also Reale, Beauregard, & Martineau, 2017) concluded that a three cluster solution identified subgroups (non-sadists, mixed group, and sadists) that differed only in severity and supported the hypothesis that sexual sadism is a dimension. The notion of sexual sadism as a dimension have been supported by several recent empirical studies on both sexual offenders and sexual murderers (e.g., Gonçalves, Rossegger, & Gerth, 2019; Longpré, Guay, & Knight, 2019a; Stefanska, Nitschke, Carter, & Mokros, 2019).

McPhail (2019a) criticized Longpré et al.'s (2018) failure to test the possibility of a three-cluster model with latent profile analysis (LPA) and noted that Mokros et al. (2014) only tested 1-class versus 2-class models with LPA, finding that both models fit the data equally well. Longpré et al. (2019c) addressed McPhail's (2019a) concern. Using the conservative Lo-Mendel-Rubin (LMR) likelihood ratio to determine the cutoff for classes in their LPT, they found that neither a 2-class nor a 3-class solution significantly improved model fit over the 1-class solution. Using a less conservative test, the bootstrap likelihood ratio, the 2-class, but not the 3-class solution reached significance. Subsequent analyses of the 2-class solution indicated that like Reale's (2017) results groups differed in severity and not in kind. Although LPA

provides a metric for determining how many subgroups might be identified in a sample, like all cluster analytic solutions it cannot address definitively the issue of the taxonicity of the boundary between groups (Ruscio et al., 2006).

Conclusion

The overwhelming consistency of results supporting a dimensional structure of sexual sadism should not be jettisoned on the basis of one ambiguous result as McPhail (2019a) suggests. Consistency across multiple procedures, measures, and samples is both a cornerstone and necessary component of a taxometric investigation that should prevent mistaken conclusions. In his final publication on taxometrics, Meehl (2004) highlighted the importance of consistency:

> A crucial feature of the coherent cut kinetics method of taxometrics is reliance on consistency tests to provide multiple lines of evidence. I have always advocated that taxometricians should use multiple taxometric procedures and consistency tests. If the latent structure is taxonic, one sort of coherent picture will emerge; if it is nontaxonic, a different sort of picture will emerge (p. 42).

Although the idea of investigating a trichotomous model of sexual sadism might seem intriguing, diverse results across multiple samples have shown conclusively that the so-called "ghost" in Longpré et al.'s (2018) study was nothing more than noise in an L-Mode curve. In line with Meehl's vision of how researchers should use taxometric procedures and on the basis of a convergence of results across procedures, measures, and samples, we believe that Longpré et al.'s (2018) conclusions were sound, and the purported trichotomous "ghost" is a snark that does not haunt their data. McPhail's mistakes provide both a cautionary tale about the pitfalls of interpreting individual figures in taxometrics and a reassertion of the importance of the fundamentals behind Meehl's analytic strategy.

References

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Press.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302. http://dx.doi.org/10.1037/h0040957

Gonçalves, L. C., Rossegger, A., & Gerth, J. (2019). Sexual sadism among sex offenders in Switzerland. *Sexual Abuse*. https://doi.org/10.1177/1079063218800473

Haslam, N., & Cleland, C. (2002). Taxometric analysis of fuzzy categories: A Monte Carlo study. *Psychological Assessment, 90*(2), 401-404. https://doi.org/10.2466/pr0.2002.90.2.401

Knight, R. A., Sims-Knight, J., & Guay, J. P. (2013). Is a separate diagnostic category defensible for paraphilic coercion? *Journal of Criminal Justice, 41,* 90-99. https://doi.org/10.1016/j.jcrimjus.2012.11.002

Lahey, B. B., & Waldman, I. D. (2003). A developmental propensity model of the origins of conduct problems during childhood and adolescence. In B. B. Lahey, T. E. Moffitt, & A. Caspi (Eds.), *Causes of conduct disorder and juvenile delinquency* (pp. 76-117). New York: The Guilford Press.

Longpré, N., Guay, J. P. & Knight, R. A. (2019a). MTC Sadism Scale: Toward dimensional assessment of severe sexual sadism. *Assessment*, *26*(1), 70-84. doi:10.1177/1073191117737377

Longpré, N., Guay, J. P., Knight, R. A., & Benbouriche, M. (2018). Sadistic offender or sexual sadism? Taxometric evidence for a dimensional structure of sexual sadism. *Archives of Sexual Behavior, 47*(2), 403-416. doi:10.1007/s10508-017-1068-4

Longpré, N., Knight, R. A. & Guay, J. P. (2019b, July). *Latent structure and covariates of the Agonistic Continuum among community sample and college students.* Oral presentation at the Australian and New Zealand Association for the Treatment of Sexual Abuse (ANZATSA) Conference, Brisbane, Australia.

Longpré, N., Sims-Knight, J. E., Neumann, C., Guay, J. P., & Knight, R. A. (2019c). *Is paraphilic coercion a different construct from sadism or simply the lower end of an Agonistic Continuum?* Manuscript submitted for publication.

McGrath, R. E. (2008). Inferential errors in taxometric analyses of ordered three-class constructs. *Journal of Personality Assessment, 90,* 11–25. doi:10.1080/00223890701356755

McGrath, R. E., & Walters, G. D. (2012). Taxometric analysis as a general strategy for distinguishing categorical from dimensional latent structure. *Psychological Methods, 17*, 284–293.

McPhail, I. V. (2019a). Of dimensions, dichotomies, and trichotomies: Comments on the taxometric ghost that haunts Longpré, Guay, Knight, and Benbouriche (2018) [Letter to the Editor]. *Archives of Sexual Behavior*. doi:10.1007/s10508-019-01530-y

McPhail, I. V. (2019b). *Conceptual and empirical issues in pedohebephilic interest*. Unpublished doctoral dissertation, University of Saskatchewan, Saskatoon, SK.

McPhail, I. V., Olver, M. E., Brouillette-Alarie, S., & Looman, J. (2018). Taxometric analysis of the latent structure of pedophilic interest. *Archives of Sexual Behavior, 47*(8), 2223-2240. doi:10.1007/s10508-018-1225-4

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry, 1,* 108-141.

Meehl, P. E. (1995). Boot straps taxometrics: Solving the classification problem in psychopathology. *American Psychologist, 50*(4), 266–275. doi:10.1037/0003-066x.50.4.266

Meehl, P. E. (2004). What's in a taxon? *Journal of Abnormal Psychology, 113,* 39–43. http://dx.doi.org/10.1037/0021-843X.113.1.39

Meehl, P. E., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports*, *74*, 1059–1274. https://doi.org/10.2466/pr0.1996.78.3c.1091

Mokros, A., Schilling, F., Weiss, K., Nitschke, J., & Eher, R. (2014). Sadism in sexual offenders: Evidence for dimensionality. *Psychological Assessment, 26*(1), 138-147. http://dx.doi.org/10.1037/a0034861

Reale, K. S. (2017). *An examination of sadism in sexual homicide: Are investigative awareness and the severity of sadistic behaviour distinctive features?* Unpublished master's thesis, Simon Fraser University, Vancouver, BC.

Reale, K. S., Beauregard, E., & Martineau, M. (2017). Sadism in sexual homicide offenders: Identifying distinct groups. *Journal of Clinical Psychology, 7*(2), 120-133. doi:10.1108/JCP-11-2016-0042

Ruscio, J. (2007). The clinician as subject: Practitioners are prone to the same judgment errors as everyone else. In S. O. Lilienfeld & W. T. O'Donohue (Eds.), *The great ideas of clinical science: 17 principles that every mental health professional should understand* (pp. 29-47). New York: Routledge/Taylor & Francis Group.

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

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*(2), 241-246. http://dx.doi.org/10.1037/1040-3590.19.2.241

Ruscio, J., & Ruscio, A. M. (2004). A nontechnical introduction to the taxometric method. *Understanding Statistics*, *3*, 151–194. doi:10.1207/s15328031us0303_2

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*(2), 349-386. doi:10.1080/00273170701360795

Ruscio, J., Walters, G. D., Markus, D. K., & Kaczetow, W. (2010). Comparing the relative fit of categorical and dimensional latent variable models using consistency tests. *Psychological Assessment*, *22*(1), 5-21. doi:10.1037/a0018259

Stefanska, E. B., Nitschke, J., Carter, A. J., & Mokros, A. (2019). Sadism among sexual homicide offenders: Validation of the Sexual Sadism Scale. *Psychological Assessment*, *31*(1), 132-137. http://dx.doi.org/10.1037/pas0000653

Walters, G. D., McGrath, R. E., & Knight, R. A. (2010). Taxometrics, polytomous constructs, and the comparison curve fit index: A Monte Carlo analysis. *Psychological Assessment*, *22*, 149–156. doi:10.1037/a0017819

Walters, G. D., & Ruscio, J. (2009). Using comparison data to differentiate categorical and dimensional data by examining factor score distributions: resolving the mode problem. *Psychological Assessment*, *21*(4), 578-594. doi:10.1037/a0016558.

Walters, G. D., & Ruscio, J. (2010). Where do we draw the line? assigning cases to subsamples for MAMBAC, MAXCOV, and MAXEIG taxometric analyses. *Assessment*, *17*(3), 321-333. doi:10.1177/1073191109356539

Walton, K. E., Ormel, J., & Krueger, R. F. (2011). The dimensional nature of externalizing behaviors in adolescence: Evidence from a direct comparison of categorical, dimensional, and hybrid models. *Journal of Abnormal Child Psychology, 39*(4), 553-561. http://dx.doi.org/10.1007/s10802-010-9478-y