The Perils of Untested Assumptions in Theory Testing: A Reply to Patrick et al. (2020)

Sandeep Roy
University of North Texas

Colin Vize
Purdue University

Kasia Uzieblo
Ghent University, Ghent
Vrije Universiteit Brussels, Brussels
De Forensische Zorgspecialisten, Utrecht

Josanne D. M. van Dongen
Erasmus University, Rotterdam

Joshua D. Miller
University of Georgia

Donald R. Lynam
Purdue University

Inti Brazil
Radboud University, Nijmegen

Dahlnym Yoon
FernUniversität in Hagen

Andreas Mokros
FernUniversität in Hagen

Nicola S. Gray
Swansea University

Robert Snowden
Cardiff University

Craig S. Neumann
University of North Texas

Author Note.
Sandeep Roy and Colin Vize are joint first authors. Correspondence: Sandeep Roy, Psychology, University of North Texas, Denton, TX. 76203 or Colin Vize, Purdue University, 703 3rd Street, West Lafayette, IN Email: SandeepRoy@my.unt.edu; cvize@purdue.edu;

In press, PDTRT

Patrick and colleagues (2020) suggested Roy et al. (2020) were misguided on statistical and conceptual grounds. We respectfully disagree and maintain Patrick et al. have not provided compelling evidence that the TriPM structure accords with the triarchic model of psychopathy. In addition to Roy et al., we discuss other studies that document the mis-specification and multidimensionality of the TriPM domains (Collison et al., 2020; Somma et al., 2019; Stanton et al., 2020). We also show their exploratory structural equation model (ESEM) is not optimal.

Inherent Multidimensionality Within Triarchic Model Framework

The goal of Roy et al. was to test the degree to which the TriPM structure aligned with the triarchic framework. We found replicable evidence for a seven-factor model and then used this model to highlight the multidimensionality of all three TriPM scales[1]. Instead of taking the *a priori* assumptions of the TriPM for granted, they were tested empirically, as is done in any scientific endeavor. The paucity of structural analyses on the TriPM is puzzling given its popularity. For instance, the second edition of the Handbook of Psychopathy (Patrick, 2018) uses the triarchic model as an organizing framework. Most measures of psychopathy have undergone intensive structural analyses (see Roy et al., 2020). It was surprising to us that despite the *many thousands* of data points available to the 22 authors of the Patrick et al. response, they chose not to demonstrate the robust structural validity of the TriPM[2]. Instead, Patrick et al. (2020) primarily focused on modeling general personality items (i.e., NEO-FFI) via ESEM and a single data set with the TriPM. The focus on the modeling of the NEO-FFI was tangential to the topic at hand (structural validity of the TriPM) and a missed opportunity to engage in comprehensive item-level analysis of the TriPM using the immense amount of data available to them.

---

[1] We also found adequate fit ((CFI = .90; RMSEA = .06) for the seven-factor model using the Stanton et al. (2020) community data (personal and data communication June 20, 2020).

[2] Using Sleep et al.'s (2020) TriPM meta-analysis as a reference and omitting Triarchic-derivative studies, we estimated Patrick and his colleagues (2020) have a combined TriPM mega-sample of 18,822 (college N = 7,957, community N = 6,807, correctional N = 1,232, adolescent N = 2,374, unknown N = 452) participants.

It is noteworthy that three recent studies involving TriPM item level analyses (9 samples in total) have all raised numerous concerns (Collison et al., 2020; Stanton et al., 2020; Roy et al., 2020) and the problems identified tend to be quite consistent irrespective of the approach used. All three studies identified similar problems such as poor statistical fit for the *a priori* triarchic model, items with poor psychometric properties (e.g., item cross-loadings, low factor loadings), and the presence of significant multidimensionality within the TriPM domains.

The multidimensionality evident in each TriPM domain is not surprising given their development. Fearless Dominance (FD), an empirical referent for Boldness (Patrick et al., 2020), is multidimensional (see Roy et al., 2020). The DSM-5 psychopathy specifier, created to instantiate FD/Boldness in the diagnostic nosology, contains multiple dimensions that have differential relationships with external correlates (Miller et al., 2018), which Roy et al. replicated when examining the TriPM Boldness scale. Stanton et al. (2020) also found Boldness to be multidimensional. Additionally, meta-analytic studies of the PPI (see Miller & Lynam, 2012) and TriPM (Sleep, Weiss, Miller, & Lynam, 2019) found that FD and Boldness are defined by multiple dimensions—i.e., low Neuroticism and high Extraversion (meta-analytic correlations of Boldness with these two domains were -.60 and .58, respectively).

One important consequence of the multidimensionality of the TriPM scales is that it leads to substantial overlap between the Meanness and Disinhibition scales. TriPM Meanness combines a variety of dissocial tendencies (callousness, exploitative, empowerment through cruelty or destructiveness) into one domain (Patrick et al., 2013). Disinhibition was designed to index both impulsivity *and* antisocial tendencies (Patrick et al., 2013). Patrick et al. (2020) state items with “the ability to differentiate disinhibitory tendencies from callous-aggressiveness” (p. 5) were selected to index the TriPM Disinhibition and Meanness scales. Yet, meta-analytic findings reveal the Meanness and Disinhibition scales evidence poor discriminant validity given their highly (absolute) similarity with regard to external correlates (*rICC* = .83; Sleep et al., 2019). Collison et al. (2020) also found little distinction between the Meanness and Disinhibition scales across Triarchic-derivative scales when examining correlations between EFA-derived factor scores with the manifest triarchic scales[3]. For example, in the three-factor solution for the TriPM, Collison et al. found one factor that corresponded well with *a priori* Boldness (convergent *r* = .97), a second factor that corresponded moderately well with *a priori* Meanness, and a third factor that was equally strongly related to both *a priori* Meanness (*r* = .86) and *a priori* Disinhibition (*r* = .95). Our findings suggest that the lack of differentiation between these TriPM scales is due to both scales tapping dissocial tendencies. Examination of item content from these domains reveals the problems, such as Collision et al. (2020, p. 8) noted, "when examining items in the Meanness domain, many of which deal with antisocial behavior and delinquency (e.g., "I would enjoy being in a high-speed chase"; "I enjoy a good physical fight"), which likely have a disinhibitory or impulsive quality to them. The same concern holds for Disinhibition items with antagonistic content (e.g., "I have stolen something out of a vehicle", "I have conned people to get money from them"; "I have robbed someone") which are likely to be multiply determined and perhaps even more strongly driven by antagonism/meanness."

Across multiple samples, Roy et al. found that the Enjoy Hurting Others factor (extracted from Meanness), and the Antisocial factor (extracted from Disinhibition), were more highly correlated with each other (*r'*s = .64 - .80) than other factors extracted from the original scales. This can also be seen in Collison et al. (2020) where, in their six-factor structure, the antisocial factor correlated equally highly with Meanness (*r* = .84) and Disinhibition (*r* = .84).

---

[3] Patrick and colleagues suggested that the use of EFA in Collison et al. (2020) was inappropriate, and that instead ESEM should have been used. However, in Mplus both EFA and ESEM routines will yield identical results for the same solution (e.g., 3 factors).

To further examine a three-factor ESEM for TriPM items, we generated factor scores with our samples, and then correlated the factor scores with the *a priori* TriPM scale scores. Strikingly, there were heterogeneous divergent associations between Meanness factor scores and *a priori* Disinhibition (*r's* = .10 to .59), suggesting item cross-loading instability across samples. More problematic, we found strong *divergent* correlations between Disinhibition factor scores with *a priori* Meanness (*r's* = .61 to .72), highlighting significant overlap between Disinhibition and Meanness items. Boldness factor scores had somewhat smaller divergent associations with the Meanness (*r's* = -.26 to .12) and Disinhibition (*r's* = -.60 to -.21) scales, and similarly for Meanness factor (*r's* = .12 to .42) and Disinhibition factor (*r's* = -.15 to .35) scores with *a priori* Boldness, but were heterogeneous and highlight again three-factor ESEM instability. In sum, the evidence does not fit with the assumption that there are three clear TriPM domains[4].

Testing Model Fit. Using the Roy et al. samples, we compared our seven-factor CFA model to the Patrick et al. target-rotated three-factor ESEM, which includes 153 correlated residual errors. As noted in Roy et al., correlated residuals are problematic, in part, because they often fail to replicate (MacCallum et al., 1992). Also, standard TriPM scoring does not account for item cross-loadings or extensive residual correlations, highlighting a disconnect between the ESEM and practical use of the measure. Still, we tested the Roy et al. seven-factor model and the Patrick et al. three-factor ESEM with and without correlated residuals. Mplus MODINICIES identified 72 residuals for our seven-factor model using sample 1. Table 1 shows little difference in model fit across samples for the CFA and ESEM, both showing adequate fit[5]. In contrast to

[4] In on-going research, we demonstrate how the use of the multidimensional triarchic scales hinder the identification of person-centered variants of psychopathy relative to Roy et al.'s seven factor model, challenging the view of the TriPM domains are optimal 'building blocks' of psychopathy.

[5] We did not employ ΔCFI to further assess fit but would urge readers to be aware that, for models that are not mis-specified, ΔCFI appears viable when using the WLSMV estimator (Sass et al., 2014). Moreover, models with the same items but different number of factors are nested, and thus there is emerging support for using ΔCFI to assess differences in model fit with WLSMV.

these results, Patrick et al. reported near perfect fit for their TriPM ESEM, which also differs substantially from the .90 TLI reported by Somma et al. (2019). Adding correlated residuals led to meaningful increases in model fit which is expected since they are akin to method factors (e.g., wording effects) embedded within the model[6]. Yet, Patrick et al. reported little difference in model fit without these residuals. Not surprisingly, 40-65% of the ESEMs correlated residuals failed to reach significance in our analyses, illustrating the sample-specific nature of these parameters. Mean ESEM item cross-loadings for Boldness (B), Meanness (M), and Disinhibition (D) were generally small (.12, .16, and .23, respectively). However, as Roy et al. found, some items had large cross-loadings (B *range* = .39 - .44; M = .32 - .69; D = .49 - .57). Patrick et al. reported much smaller cross-loadings than we find across six samples, as well as those reported by Somma et al. Taken together, the generalizability of the Patrick et al. results cannot be assumed given they are based on a singular data set and are inconsistent with our results across six samples, as well as those reported by Somma et al.[7] Alas, Patrick and colleagues' failure to further test their three- or our seven-factor model via their extensive TriPM datasets is a missed opportunity.

Limited Utility of Broad Factors in Advancing Theory Development

Patrick and colleagues (2020) argue that the nomological network should be considered to assess utility of the TriPM. However, given the multidimensionality of the domains, one cannot be certain as to why the TriPM scales show observed relations with relevant criteria (Strauss & Smith, 2009). Based on their ESEM, associations between the TriPM factors and external correlates may be attributable to items with large cross-loadings, the covariance among 153 correlated residuals, or the items within a target factor. Given this level of psychometric

[6] The Patrick et al. ESEM requires 388 parameters to account for the covariance of 58 items (58 loadings for each factor, 58 error variances, 3 factor correlations, and 153 correlated residuals), while our CFA requires 101 parameters to account for 40 items (40 loadings, 40 error variances, 21 factor correlations). Thus, the ESEM is far more saturated and a much easier model to fit (data points: parameters = 4.4, compared to our CFA = 8.2).

[7] Patrick et al. (2020) provide little detail regarding their replication sample, other than it was essentially a young college sample primarily composed of females.

uncertainty, it is questionable whether the TriPM scales can help advance theory and clinical practice. In contrast, the domains of the FFM have demonstrated theoretical utility in advancing understanding of personality, due in part, to lower-order *unidimensional* facets that underlie the domains. The FFM facets improve prediction of relevant criteria compared to the broad FFM factors (Soto & John, 2017; Straus & Smith, 2009). Vize et al. (2018) found specific facets of Neuroticism (e.g., hostility, impulsivity) and Extraversion (e.g., excitement-seeking, low warmth) were more robustly associated with externalizing behaviors, compared to their respective broad domains. Similarly, Watson and colleagues (2015) showed that specific features of Extraversion which underlie the broader domain (e.g., Positive Emotionality, Sociability) bear divergent relations with various measures of psychopathology.

To examine the utility of the Patrick et al. ESEM, versus our seven-factor model, we ran structural equation models (SEMs), using the samples in Roy et al. with external correlates. Factors were allowed to correlate to account for their shared variance. As shown in Table 2, the seven-factor SEMs, despite using fewer items, accounted for more variance than the three-factor ESEMs. The results for the seven-factor model also revealed meaningful divergent associations with the external correlates (e.g., Leader vs. Stress Immunity) and associations that may advance theory (Enjoy Hurting with positive affect). These results challenge Patrick and colleagues (2020) contention that the seven-factor model only captures item wording effects or response polarity. Stanton et al. (2020) similarly demonstrated that a subset of 35 TriPM items were able to account for a range of external correlates. Taken together, our results and those from Stanton et al., demonstrate that unidimensional factors, based on fewer items than the original TriPM scales, are far from ‘wording’ or ‘method’ factors, and instead offer coherent item-to-factor domains that can help advance theory testing on larger constructs (Strauss & Smith, 2009).

Conclusions. Patrick et al. relied on a single sample of college students with TriPM data and presented results inconsistent with our analyses of general population samples, and some of their own previously published TriPM results (Somma et al. 2019). Based on meta-analytic and structural research of the TriPM (Collison et al., 2020; Miller at al., 2018; Roy et al., 2020; Sleep et al., 2019; Stanton et al., 2020), there is clear evidence of multidimensionality for each triarchic scale and continued concerns regarding the structural validity of the TriPM. Moreover, the original TriPM scales used in contemporary research do not match up with the ESEM model that Patrick et al. employ with item cross-loadings (some very large) and 153 correlated residuals. The six- (Collison et al., 2020) and seven-factor (Roy et al., 2020) models, while not ground truth per se, nevertheless provide opportunities for more dynamic explorations of the psychopathy construct.

The viability of any given model depends on the quality of its indicators and guiding theory that go into it (Roy et al., 2020). In the end, statistical models are tools that help us understand larger constructs, much like maps help us understand the vast territories they represent. Although ESEM-derived personality models may be viable (if not composed of items with large cross-loadings or residual correlations like the TriPM), CFA-derived models are also useful because they provide clear and unambiguous factors (Yue et al., 2019). CFA-based models have been helpful for representing major domains of psychopathology (Krueger & Markon, 2006), as well as general (Soto & John, 2012) and pathological personality (Neumann et al. 2015). Our results and other findings suggest the original three triarchic domains may be too broad to capture the complexity of psychopathic personality and individual variation therein. Dr. Patrick and his colleagues research have no doubt furthered understanding of psychopathy. Still, to continue to move the field forward it is important to evaluate and adjust our prevailing assumptions in the light of novel empirical findings with the goal of furthering understanding of a construct that has huge impact in society and the criminal justice system.

References

Collison, K. L., Miller, J. D., & Lynam, D. R. (2020). Examining the factor structure and validity of the Triarchic Model of Psychopathy across measures. *Personality Disorders: Theory, Research, and Treatment*. Advance online publication. doi: 10.1037/per0000394

Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review Clinical Psychology, 2*, 111-133. doi: 10.1146/annurev.clinpsy.2.022305.095213

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological Bulletin*, *111*, 490-504. doi: 10.1037/0033-2909.111.3.490

Miller, J. D., Lamkin, J., Maples-Keller, J. L., Sleep, C. E., & Lynam, D. R. (2018). A test of the empirical profile and coherence of the DSM–5 psychopathy specifier. *Psychological Assessment*, *30*(7), 870-881. doi: 10.1037/pas0000536

Miller, J. D., & Lynam, D. R. (2012). An examination of the Psychopathic Personality Inventory's nomological network: A meta-analytic review. *Personality Disorders: Theory, Research, and Treatment*, *3*(3), 305-326. doi: 10.1037/a0024567

Neumann, C. S., Hare, R. D., & Pardini, D. A. (2015). Antisociality and the construct of psychopathy: Data from across the globe. *Journal of Personality, 8*3, 678–692. doi: 10.1111/jopy.12127

Patrick, C. J. (Ed.). (2018). *Handbook of psychopathy*. Guilford Publications.

Patrick, C. J., Kramer, M. D., Krueger, R. F., & Markon, K. E. (2013). Optimizing efficiency of psychopathology assessment through quantitative modeling: Development of a brief form of the Externalizing Spectrum Inventory. *Psychological Assessment, 25*(4), 1332–1348. doi: 10.1037/a0034864

Patrick, C. J., Joyner, K. J., Watts, A. L., Lilienfeld, S. O., Somma, A., Fossati, A. et al. Latent Variable Modeling of Item-Based Factor Scales: Comment on *Triarchic or Septarchic?—Uncovering the Triarchic Psychopathy Measure's (TriPM) Structure*, by Roy et al. *Personality Disorders: Theory, Research, and Treatment*

Roy, S., Vize, C., Uzieblo, K., van Dongen, J. D. M., Miller, J., Lynam, D., et al. & Neumann, C. S. (2020, January 23). Triarchic or Septarchic?—Uncovering the Triarchic Psychopathy Measure's (TriPM) Structure. *Personality Disorders: Theory, Research, and Treatment*. Advance online publication. doi: 10.1037/per0000392

Sass, D.A., Schmitt, T.A. & Marsh, H.W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling, 21*(2), 167-180. doi: 10.1080/10705511.2014.882658

Sleep, C. E., Weiss, B., Lynam, D. R., & Miller, J. D. (2019). An examination of the triarchic model of psychopathy's nomological network: A meta-analytic review. *Clinical Psychology Review, 71*, 1–26. doi: 10.1016/j.cpr.2019.04.005

Somma, A., Borroni, S., Drislane, L. E., Patrick, C. J., & Fossati, A. (2019). Modeling the structure of the triarchic psychopathy measure: Conceptual, empirical, and analytic considerations. *Journal of personality disorders, 33*(4), 470-496. doi: 10.1521/pedi_2018_32_354

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, *113*(1), 117-143. doi: 10.1037/pspp0000096

Stanton, K., Brown, M. F., & Watson, D. (2020). Examining the Item-Level Structure of the Triarchic Psychopathy Measure: Sharpening Assessment of Psychopathy Constructs. *Assessment*, online first doi: 1073191120927786.

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, *5*, 1-25. doi: 10.1146/annurev.clinpsy.032408.153639

Vize, C. E., Miller, J. D., & Lynam, D. R. (2018). FFM facets and their relations with different forms of antisocial behavior: An expanded meta-analysis. *Journal of Criminal Justice*, *57*, 67-75. doi: 10.1016/j.jcrimjus.2018.04.004

Watson, D., Stasik, S. M., Ellickson-Larew, S., & Stanton, K. (2015). Extraversion and psychopathology: A facet-level analysis. *Journal of Abnormal Psychology*, *124*(2), 432-446. doi: 10.1037/abn0000051

Yue, X., Hongyun, L., & Kit-Tai, H. (2019). A comparison of CFA, ESEM, and BSEM in test structure analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(2), 665-677. doi: 10.1080/10705511.2018.1562928

Table 1

Model Fit Results: three-factor ESEM vs. seven-factor CFA

	No correlated errors		Correlated errors*	
Sample	**CFI**	**RMSEA**	**CFI**	**RMSEA**
	3-factor ESEM			
Sample 1	.92	.05	.95	.04
Sample 2	.90	.05	.94	.04
Sample 3	.90	.05	.93	.03
Sample 4	.90	.03	.93	.03
Sample 5	.87	.04	.89	.05
Sample 6	.91	.03	.93	.03
	7-factor CFA			
Sample 1	.92	.04	.97	.04
Sample 2	.90	.06	.95	.04
Sample 3	.90	.06	.95	.04
Sample 4	.86	.04	.92	.04
Sample 5	.86	.05	.90	.04
Sample 6	.90	.06	.92	.05

Note. *The Patrick et al./Somma et al. ESEM requires 153 correlated residual errors to boost model fit, while the 7-factor CFA model only required 72 correlated errors for such a boost.

Table 2

Structural equation modeling (SEM) results: seven-factor CFA vs. three-factor ESEM

Sample 1

7-factor CFA model				**3-factor ESEM model**			
Predictor	**Criteria**			**Predictor**	**Criteria**		
	NA	***PA***	***AUDIT***		***NA***	***PA***	***AUDIT***
Lead	.35			Boldness	-.41	.56	.10
Stress Immunity	-.27	-.11					
Pos. Self	-.63	.68					
Callous		-.26		Meanness	-.07	-.28	
Enjoy Hurting		.28					
Impulsive			.16				
Antisocial			.33	Disinhibition	.21	-.08	.36
R^2	**.46**	**.48**	**.17**		**.37**	**.38**	**.15**

Sample 2

	A	*C*		*A*	*C*
Lead	-.24	.26	Boldness	-.10	.38
Stress Immunity	.21				
Pos. Self					
Callous	-.58		Meanness	-.71	-.16
Enjoy Hurting	-.26				
Impulsive		-.76			
Antisocial	-.15		Disinhibition	-.17	-.56
R^2	**.70**	**.68**		**.64**	**.59**

Sample 3

	ASB		*ASB*
Lead		Boldness	.21
Stress Immunity	.46		
Pos. Self			
Callous		Meanness	
Enjoy Hurting			
Impulsive			
Antisocial	.43	Disinhibition	.49
R^2	**.34**		**.29**

Note. NA = Negative Affect; PA = Positive Affective; AUDIT = Alcohol use inventory test; A = Agreeableness; C = Conscientiousness; ASB = Antisocial behavior. All *p's* < .05 - .001