Supplementary Material

**Reporting Metric of Individual Articles**

In some cases, it was unclear what the reporting metric was, so we contacted each of the authors to resolve the metric. We obtained clarification from a majority of authors. When we could not obtain clarification and could not infer the parameterization from the text or estimates, we included them in the analysis as reported in the original paper. Prior to analysis we transformed all confirmed estimates into IRT parameterized estimates if not already in that form.

**Inclusion of Individual Articles in Analysis**

In two cases, multiple publications (Bond et al., 2012; Borges et al., 2010, 2011; Cherpitel et al., 2010; Mewton et al., 2011a; Proudfoot et al. 2006), used the exact same overall sample and reported partially overlapping results. To avoid double-counting we only included the one with more complete criteria information in the analysis (Cherpitel et al., 2010; Mewton et al., 2011a). Bond and colleagues (2012) do report analyses according to the ICD-10 (WHO, 1992) diagnostic system; however, their analysis is subsumed within those of their other publications where they match items according to the DSM-IV/5 criteria sets (Borges et al., 2010, 2011; Cherpitel et al., 2010). We do not include this analysis because it uses fewer of the analyzed criteria than its sister publications, but also because it reduces the possibility of variability due to idiosyncratic criteria set differences across diagnostic systems.

Six papers reported results broken down by subgroups (Harford et al., 2009; Cherpitel et al. 2010; Duncan et al., 2011; McCutcheon, et al., 2011; Srisurapamont et al., 2012; Pruess et al., 2014). In those cases, we entered each subgroup’s severity data as separate estimates and again, to avoid double-counting, did not include the overall sample severity estimates if available. There were four groups of papers (Table 1) that reported on different but partially overlapping subsamples. They were included in the analyses, and adjusted for by estimating a variance component for unique samples.

**Inclusion of Different Measurement Models in Analysis**

 We excluded one parameter (Rasch) models from inclusion in the meta-analysis because they assume equal discrimination parameters across criteria, which can affect the estimation and relative ordering of criterion severities. Additionally, we omitted studies that report basic prevalence rates as indicators of criterion severity because they ignore the covariances between criteria. Although relative rankings between prevalence and severity as measured by IRT thresholds are maintained a majority of the time, the relative differences in them on a constant scale may change as a function of the multivariate covariances among criteria.

**Additional Interrater Reliability Considerations**

 Rater reliability was assessed after all of the coding was completed. To control for potential rater drift over the coding process, raters were trained using a highly structured set of coding guidelines for each variable that minimized the amount of subjective bias that could be introduced. In addition, a vast majority of the information coded were concrete numeric values that likely minimized subjective interpretation. The coders were also not aware of the study hypotheses. Raters did code the manuscripts in a predefined order, and so drift could be provisionally assessed by splitting the data approximately in half, corresponding to earlier and later ratings, and calculating whether reliabilities between raters differed from the first and second halves. As in the entire data set, the reliabilities in the first (*ICC*s = .97 to 1.00, κs = 1.00) and second (*ICC*s = 1.00, κs = .68 to 1.00) halves were good to near perfect. Only one variable (out of 31) exhibited a significant difference from earlier and later ratings (clinical sample composition, κ1 = 1.00, κ2 = .68, *p* < .05). This was due to the raters not agreeing on clinical versus population-based composition for 3 cases in the second half, which in the overall data reconciliation meeting with the first author were resolved.

**Individual Bivariate Correlations between Included Articles**

Table S1 presents the individual Pearson and Spearman correlation coefficients between each of the included 30 published manuscripts included in the meta-analysis. For parsimony we averaged the severities for the manuscripts that reported results on multiple groups.

**Parallel Results for Standardized and Weighted Analyses**

Table S2 shows parallel variance component estimates to Table 2, Model 2 for analyses conducted on the within-study standardized severity estimates and weighting estimates according to each samples’ sample size relative to the total number of individuals across all included studies. By standardizing we note that cross-sample variance in overall severity level is removed so any sample level variance (i.e. main effects of instrument, timeframe, clinical population) will necessarily be zero.

**Consideration of Criteria Discrimination Parameters**

 We are less interested in the reliability of discriminations across studies as they confer information about sensitivity of criteria at a given level of latent severity. The utility of this information, is predicated on where the inflection point of the IRT information curve is located (i.e. what the latent severity/threshold is). Rather, from a practical standpoint, we are most interested in if criteria are systematically more or less difficult to endorse for individuals, and if not, what factors might account for such inconsistency.

As a preliminary investigation we replicated our analysis of the IRT severity parameters on the parallel discrimination parameters. We observed a nonsignificant amount of the overall variance as explainable by systematic criteria discrimination values across studies (σ2 = .03, *SE* = .03, *p* = .188), which resulted in an overall reliability estimate of .09. However, as with the severity parameters, there was a highly significant criterion-by-instrument variance (σ2 = .09, *SE* = .03, *p* < .001). This result suggests that many differential item functioning (DIF) results may not be generalizable beyond the instrument used to assess the criteria, and even within a given instrument the reliability of the DIF is .39.

Additional References

**Beseler CL, Taylor LA, & Leeman RF** (2010). An item-response theory analysis of DSM-IV alcohol-use disorder criteria and “binge” drinking in undergraduates. *Journal of Studies on Alcohol and Drugs* **71**, 418–423.

**Bond J, Ye Y, Cherpitel CJ, Borges G, Cremonte M, Moskalewicz J, & Swiatkiewicz G** (2012). Scaling properties of the combined ICD-10 dependence and harms criteria and comparisons with DSM-5 alcohol use disorder criteria among patients in the emergency department. *Journal of Studies on Alcohol and Drugs* **73**, 328-336.

**Borges G, Cherpitel CJ, Ye Y, Bond J, Cremonte M, Moskalewicz J, & Swiatkiewicz G** (2011). Threshold and optimal cut-points for alcohol use disorders among patients in the emergency department. *Alcoholism: Clinical & Experimental Research* **35**, 1270–1276.

**Derringer J, Krueger RF, Dick DM, Agrawal A, Bucholz KK, Foroud T, Grucza RA, Hesselbrock MN, Hesselbrock V, Kramer J, Nurnberger Jr JI, Schuckit M, Bierut LJ, Iacono WG, & McGue M** (2013). Measurement invariance of DSM‐IV alcohol, marijuana and cocaine dependence between community‐sampled and clinically over-selected studies. *Addiction* **108**, 1767-1776.

**Duncan AE, Agrawal A, Bucholz KK, Sartor CE, Madden PA, & Heath AC** (2011). Deconstructing the architecture of alcohol abuse and dependence symptoms in a community sample of late adolescent and emerging adult women: An item response approach. *Drug and Alcohol Dependence* **116**, 222-227.

**Edwards AC, Gillespie NA, Aggen SH, & Kendler KS** (2013). Assessment of a modified DSM‐5 diagnosis of alcohol use disorder in a genetically informative population. *Alcoholism: Clinical and Experimental Research* **37**, 443-451.

**Ehlke SJ, Hagman BT, & Cohn AM** (2012). Modeling the dimensionality of DSM-IV alcohol use disorder criteria in a nationally representative sample of college students. *Substance Use & Misuse* **47**, 1073-1085.

**Gelhorn H, Hartman C, Sakai J, Stallings M, Young S, Rhee S, Corley R, Hewitt J, Hopfer C, & Crowley T** (2008). Toward DSM-V: An item response theory analysis of the diagnostic process for DSM-IV alcohol abuse and dependence in adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry* **47**, 1329–1339.

**Gilder DA, Gizer IR, & Ehlers CL** (2011). Item response theory analysis of binge drinking and its relationship to lifetime alcohol use disorder symptom severity in an American Indian community sample. *Alcoholism: Clinical & Experimental Research* **35**, 984–995.

**Hagman BT, & Cohn AM** (2011). Toward DSM-V: Mapping the alcohol use disorder continuum in college students. *Drug and Alcohol Dependence* **118**, 202–208.

**Hasin DS, Fenton MC, Beseler C, Park JY, & Wall MM** (2012). Analyses related to the development of DSM-5 criteria for substance use related disorders: 2. Proposed DSM-5 criteria for alcohol, cannabis, cocaine and heroin disorders in 663 substance abuse patients. *Drug and Alcohol Dependence* **122**, 28–37.

**Keyes KM, Krueger RF, Grant BF, & Hasin DS** (2011). Alcohol craving and the dimensionality of alcohol disorders. *Psychological Medicine* **41**, 629–640.

**Kuerbis AN, Hagman BT, & Morgenstern J** (2013a). Alcohol Use Disorders among Substance Dependent Women on Temporary Assistance with Needy Families: More Information for Diagnostic Modifications for DSM‐5. *The American Journal on Addictions* **22**, 402-410.

**Kuerbis AN, Hagman BT, & Sacco P** (2013b). Functioning of alcohol use disorders criteria among middle-aged and older adults: Implications for DSM-5.*Substance Use & Misuse* **48**, 309-322.

**McCutcheon VV, Agrawal A, Heath AC, Edenberg HJ, Hesselbrock VM, Schuckit MA, Kramer JR, & Bucholz KK** (2011). Functioning of alcohol use disorder criteria among men and women with arrests for driving under the influence of alcohol. *Alcoholism: Clinical & Experimental Research* **35**, 1985–1993.

**Mewton L, Slade T, McBride O, Grove R, & Teesson M** (2011a). An evaluation of the proposed SM-5 alcohol use disorder criteria using Australian national data. *Addiction* **106**, 941–950.

**Mewton L, Teesson M, Slade T, & Cottler L** (2011b). Psychometric performance of DSM-IV alcohol use disorders in young adulthood: Evidence from an Australian general population sample. *Journal of Studies on Alcohol and Drugs* **72**, 811–822.

**Preuss UW, Watzke S, & Wurst FM** (2014). Dimensionality and stages of severity of DSM-5 criteria in an international sample of alcohol-consuming individuals. *Psychological Medicine* **44**, 3303-3314.

**Proudfoot H, Baillie AJ, & Teeson M** (2006). The structure of alcohol dependence in the community. *Drug and Alcohol Dependence* **81**, 21-26.

**Rose JS, Lee CT, Selya AS, & Dierker LC** (2012). DSM-IV alcohol abuse and dependence criteria characteristics for recent onset adolescent drinkers. *Drug and Alcohol Dependence* **124**, 88-94.

**Saha TD, Chou SP, & Grant BF** (2006). Toward an alcohol use disorder continuum using item response theory: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Psychological Medicine* **36**, 931–941.

**Saha TD, Stinson FS, & Grant BF** (2007). The role of alcohol consumption in future classifications of alcohol use disorders. *Drug and Alcohol Dependence* **89**, 82–92.

**Wu LT, Pan JJ, Blazer DG, Tau B, Stitzer ML, Brooner RK, Woody GE, Patkar AA, Blaine JD** (2009). An item response theory modeling of alcohol and marijuana dependences: A National Drug Abuse Treatment Clinical Trials Network study. *Journal of Studies on Alcohol and Drugs* **70**, 414–425.

**Wu LT, Blazer DG, Woody GE, Burchett B, Yang C, Pan JJ, & Ling W** (2012). Alcohol and drug dependence symptom items as brief screeners for substance use disorders: Results from the Clinical Trials Network. *Journal of Psychiatric Research* **46**, 360-369.

Table S1. Bivariate Pearson and Spearman correlations between IRT study severity parameters.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Article | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
| 1. | Casey et al., 2012 | 1.00 | .93 | .93 | .89 | .61 | .72 | .65 | .47 |
| 2. | Dawson et al., 2010 | .83 | 1.00 | 1.00 | .99 | .87 | .85 | .83 | .60 |
| 3. | Saha et al., 2006 | .84 | 1.00 | 1.00 | .98 | .83 | .79 | .83 | .60 |
| 4. | Saha et al., 2007 | .83 | 1.00 | 1.00 | 1.00 | .87 | .81 | .81 | .53 |
| 5. | Schmulewitz et al., 2010a | .43 | .86 | .80 | .85 | 1.00 | .68 | .73 | .60 |
| 6. | Keyes et al., 2011 | .81 | .85 | .78 | .85 | .75 | 1.00 | .75 | .16 |
| 7. | Preuss et al., 2014 | .56 | .70 | .68 | .70 | .54 | .70 | 1.00 | .41 |
| 8. | Mewton et al., 2011a | .59 | .65 | .67 | .65 | .71 | .44 | .47 | 1.00 |
| 9. | Mewton et al., 2011b | .47 | .72 | .66 | .72 | .81 | .60 | .48 | .90 |
| 10. | McCutcheon et al., 2011b | .39 | .50 | .49 | .51 | .62 | .59 | .68 | .55 |
| 11. | Beseler et al., 2010 | -.04 | .28 | .25 | .28 | .57 | .10 | .12 | .55 |
| 12. | Hagman & Cohn, 2011 | .01 | .49 | .30 | .49 | .79 | .36 | .01 | .47 |
| 13. | Ehlke et al., 2012 | -.47 | .12 | -.14 | .12 | .39 | -.08 | -.38 | -.17 |
| 14. | Kuerbis et al., 2013b | -.19 | .40 | -.04 | .39 | .61 | .30 | -.46 | .32 |
| 15. | Hagman & Cohn, 2013 | -.48 | -.26 | -.39 | -.25 | -.13 | -.31 | -.72 | -.56 |
| 16. | Rose et al., 2012 | -.70 | -.42 | -.49 | -.42 | -.23 | -.40 | -.37 | -.73 |
| 17. | Harford et al., 2009b | -.49 | .09 | -.18 | .09 | .36 | -.11 | -.47 | -.23 |
| 18. | Srisurapamont et al., 2012 | .25 | .52 | .51 | .53 | .67 | .41 | .66 | .51 |
| 19. | Duncan et al., 2011 | -.22 | .26 | .25 | .27 | .57 | -.13 | .14 | .14 |
| 20. | Derringer et al., 2013 | .33 | .53 | .52 | .53 | .73 | .40 | .47 | .81 |
| 21. | Gilder et al., 2011 | .00 | .22 | .20 | .23 | .41 | .20 | .43 | .06 |
| 22. | Gelhorn et al., 2008 | -.24 | .16 | .05 | .16 | .46 | .09 | -.06 | -.07 |
| 23. | Cherpitel et al., 2010b | .43 | .24 | .27 | .22 | .24 | .35 | .01 | .50 |
| 24. | Hasin et al., 2012 | .01 | .01 | .03 | .00 | .29 | -.15 | -.01 | .67 |
| 25. | Langenbucher et al. 2004 | .64 | .63 | .64 | .63 | .58 | .33 | .10 | .33 |
| 26. | Wu et al., 2009 | .13 | .29 | .26 | .29 | .32 | .06 | .49 | .58 |
| 27. | Wu et al., 2012 | .42 | .46 | .45 | .46 | .51 | .48 | .62 | .72 |
| 28. | Martin et al., 2006 | -.46 | -.14 | -.30 | -.14 | .25 | -.04 | -.19 | -.10 |
| 29. | Edwards et al., 2013 | .19 | .41 | .34 | .42 | .50 | .50 | .40 | .33 |
| 30. | Kuerbis et al., 2013a | .34 | .53 | .43 | .53 | .59 | .33 | -.02 | .77 |
|  | Average *r* | .18 | .41 | .34 | .41 | .51 | .31 | .22 | .36 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 9. | 10. | 11. | 12. | 13. | 14. | 15. | 16. | 17. | 18. | 19. | 20. |
| 1. | .53 | .35 | .07 | -.05 | -.53 | -.20 | -.53 | -.5 | -.45 | .29 | -.04 | .32 |
| 2. | .74 | .52 | .24 | .29 | -.10 | .17 | -.39 | -.27 | -.04 | .57 | .19 | .54 |
| 3. | .68 | .52 | .20 | .05 | -.38 | -.10 | -.53 | -.41 | -.38 | .57 | .19 | .54 |
| 4. | .74 | .56 | .26 | .34 | .00 | .18 | -.30 | -.24 | .06 | .65 | .29 | .63 |
| 5. | .71 | .53 | .36 | .46 | .01 | .23 | -.52 | .02 | .07 | .82 | .24 | .57 |
| 6. | .41 | .62 | .00 | .08 | -.38 | -.20 | -.53 | -.14 | -.33 | .50 | -.20 | .46 |
| 7. | .53 | .80 | .09 | -.08 | -.38 | -.32 | -.65 | -.27 | -.54 | .68 | .20 | .68 |
| 8. | .88 | .52 | .61 | .31 | -.18 | .56 | -.52 | -.52 | -.22 | .71 | .15 | .86 |
| 9. | 1.00 | .58 | .74 | .49 | .26 | .65 | -.22 | -.37 | .20 | .82 | .39 | .89 |
| 10. | .50 | 1.00 | .40 | .43 | -.23 | -.22 | -.52 | -.45 | -.43 | .79 | .10 | .93 |
| 11. | .70 | .53 | 1.00 | .66 | .41 | .61 | .03 | -.19 | .31 | .61 | .38 | .71 |
| 12. | .68 | .59 | .83 | 1.00 | .56 | .55 | .27 | -.07 | .62 | .54 | .36 | .64 |
| 13. | .26 | -.14 | .56 | .64 | 1.00 | .66 | .93 | .23 | .94 | .07 | .71 | .18 |
| 14. | .54 | -.32 | .46 | .63 | .70 | 1.00 | .30 | -.15 | .66 | -.11 | .38 | .00 |
| 15. | -.26 | -.61 | .13 | .21 | .89 | .42 | 1.00 | .37 | .88 | -.54 | .60 | -.31 |
| 16. | -.53 | -.57 | -.29 | -.26 | .28 | -.09 | .52 | 1.00 | .28 | -.18 | .05 | -.57 |
| 17. | .20 | -.26 | .49 | .59 | .99 | .70 | .93 | .35 | 1.00 | -.18 | .59 | -.11 |
| 18. | .61 | .75 | .40 | .39 | -.26 | -.52 | -.77 | -.34 | -.33 | 1.00 | .14 | .82 |
| 19. | .45 | .19 | .56 | .59 | .77 | .36 | .56 | .23 | .72 | .20 | 1.00 | .36 |
| 20. | .92 | .91 | .86 | .84 | .19 | .09 | -.32 | -.65 | .07 | .66 | .39 | 1.00 |
| 21. | .17 | .79 | .53 | .59 | .35 | -.16 | -.03 | -.13 | .24 | .5 | .42 | .88 |
| 22. | .24 | .40 | .71 | .73 | .61 | .25 | .47 | .17 | .59 | .37 | .47 | .66 |
| 23. | .28 | .09 | .15 | .09 | -.48 | .12 | -.36 | -.36 | -.43 | .17 | -.54 | .16 |
| 24. | .47 | .69 | .47 | .42 | -.20 | .35 | -.51 | -.51 | -.24 | .53 | -.15 | .85 |
| 25. | .21 | .28 | .21 | .43 | -.07 | .12 | -.11 | -.47 | -.04 | .41 | -.21 | .48 |
| 26. | .55 | .69 | .57 | .46 | .18 | .20 | -.41 | -.65 | .05 | .41 | .27 | .69 |
| 27. | .72 | .92 | .50 | .48 | -.25 | -.36 | -.68 | -.77 | -.36 | .82 | .06 | .85 |
| 28. | .09 | .45 | .63 | .68 | .54 | .23 | .35 | .12 | .49 | .14 | .33 | .67 |
| 29. | .43 | .57 | .42 | .52 | .30 | .17 | .01 | -.08 | .21 | .04 | .45 | .48 |
| 30. | .69 | .04 | .34 | .48 | .13 | .64 | -.22 | -.47 | .14 | .25 | .12 | .76 |
|  | .46 | .37 | .41 | .47 | .19 | .20 | -.10 | -.28 | .14 | .29 | .26 | .51 |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 21. | 22. | 23. | 24. | 25. | 26. | 27. | 28. | 29. | 30. | Average *ρ* |
| 1. | .03 | -.22 | .45 | .21 | .46 | .14 | .54 | -.48 | .39 | .44 | .22 |
| 2. | .15 | .04 | .31 | .27 | .46 | .32 | .71 | -.23 | .61 | .64 | .41 |
| 3. | .15 | -.08 | .31 | .27 | .46 | .32 | .71 | -.38 | .55 | .54 | .34 |
| 4. | .24 | .13 | .23 | .20 | .46 | .41 | .77 | -.16 | .61 | .59 | .43 |
| 5. | .27 | .27 | .27 | .30 | .52 | .32 | .57 | -.02 | .63 | .64 | .43 |
| 6. | .19 | .04 | .43 | -.02 | .29 | .29 | .68 | -.15 | .49 | .34 | .26 |
| 7. | .37 | .07 | -.04 | .27 | .08 | .64 | .86 | -.16 | .56 | .31 | .30 |
| 8. | -.16 | -.13 | .38 | .81 | .32 | .68 | .86 | -.22 | .50 | .93 | .36 |
| 9. | .18 | .20 | .30 | .61 | .35 | .64 | .82 | -.01 | .60 | .78 | .49 |
| 10. | .70 | .40 | .03 | .71 | .29 | .75 | 1.00 | .35 | .67 | .50 | .39 |
| 11. | .41 | .56 | .30 | .50 | .38 | .43 | .43 | .47 | .58 | .56 | .38 |
| 12. | .56 | .69 | .16 | .47 | .63 | .29 | .32 | .65 | .56 | .62 | .39 |
| 13. | .30 | .48 | -.56 | -.12 | -.31 | .14 | -.18 | .55 | .19 | .09 | .12 |
| 14. | -.22 | .05 | .02 | .49 | .00 | .18 | -.32 | .11 | .22 | .65 | .17 |
| 15. | .17 | .35 | -.48 | -.44 | -.33 | -.37 | -.54 | .52 | -.14 | -.30 | -.13 |
| 16. | .04 | .27 | -.16 | -.66 | -.35 | -.61 | -.68 | .32 | -.07 | -.48 | -.20 |
| 17. | .20 | .42 | -.39 | -.21 | -.05 | -.21 | -.46 | .45 | .12 | .15 | .07 |
| 18. | .68 | .64 | .25 | .54 | .76 | .64 | .79 | .43 | .50 | .54 | .46 |
| 19. | .35 | .31 | -.49 | .04 | -.41 | .07 | .07 | .30 | .54 | .20 | .21 |
| 20. | .86 | .57 | .14 | .86 | .46 | .75 | .93 | .61 | .68 | .86 | .51 |
| 21. | 1.00 | .82 | -.26 | .07 | .23 | .68 | .68 | .75 | .61 | -.04 | .31 |
| 22. | .69 | 1.00 | -.05 | .00 | .33 | .18 | .29 | .85 | .42 | -.01 | .27 |
| 23. | -.31 | -.03 | 1.00 | .14 | .70 | -.18 | .21 | -.15 | .09 | .43 | .08 |
| 24. | -.01 | -.05 | .50 | 1.00 | .25 | .82 | .89 | .09 | .50 | .86 | .30 |
| 25. | .14 | .23 | .47 | .13 | 1.00 | .39 | .46 | -.03 | .14 | .43 | .25 |
| 26. | .80 | .20 | -.26 | .86 | .29 | 1.00 | .75 | .21 | .32 | .64 | .33 |
| 27. | .66 | .27 | .14 | .89 | .33 | .69 | 1.00 | .29 | .61 | .79 | .44 |
| 28. | .72 | .81 | -.14 | .20 | -.13 | .38 | .30 | 1.00 | .45 | -.04 | .18 |
| 29. | .64 | .39 | -.16 | .04 | -.14 | .24 | .21 | .53 | 1.00 | .59 | .43 |
| 30. | -.36 | -.16 | .46 | .67 | .39 | .66 | .54 | -.19 | .05 | 1.00 | .42 |
|  | .31 | .30 | .06 | .21 | .24 | .32 | .35 | .21 | .29 | .29 |  |

*Note*. Pearson correlations on lower diagonal, Spearman correlations on upper diagonal.

a Only the correlations for current AUD symptoms are listed.

b Severities were averaged across subsamples before correlating with the other published studies.

Table S2. Variance component estimates for Model 2 using standardized, unstandardized and weighted, and standardized and weighted severity estimates.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Unstandardized, Weighted | Standardizedb | Standardized, Weightedb |
| *Parameter* | *Estimate* | *SE* | *p-value* | *Estimate* | *SE* | *p-value* | *Estimate* | *SE* | *p-value* |
| σ2CRITERION | .078 | .057 | .087 | .188 | .144 | .095 | .194 | .153 | .102 |
| σ2SAMPLE | .034 | .020 | .048 |  |  |  |  |  |  |
| σ2INSTRUMENT | .069 | .103 | .252 |  |  |  |  |  |  |
| σ2TIMEFRAME | .200 | .357 | .288 |  |  |  |  |  |  |
| σ2CLINICAL | .392 | .486 | .211 |  |  |  |  |  |  |
| σ2GENDER | .005 | .014 | .352 |  |  |  |  |  |  |
| σ2AGE | .00a | – | – |  |  |  |  |  |  |
| σ2GROUP | .402 | .132 | .001 |  |  |  |  |  |  |
| σ2CRITERION\*INSTRUMENT | .089 | .024 | <.001 | .244 | .057 | <.001 | .253 | .060 | <.001 |
| σ2CRITERION\*TIMEFRAME | .014 | .016 | .187 | .071 | .044 | .055 | .072 | .050 | .074 |
| σ2CRITERION\*CLINICAL | .014 | .011 | .103 | .043 | .026 | .051 | .049 | .030 | .053 |
| σ2CRITERION\*GENDER | .005 | .005 | .162 | .009 | .013 | .255 | .012 | .011 | .132 |
| σ2CRITERION\*AGE | .075 | .021 | <.001 | .138 | .042 | <.001 | .154 | .043 | <.001 |
| σ2ERROR | .002 | .000 | <.001 | .250 | .020 | < .001 | .004 | .000 | <.001 |
|  |  |  |  |  |  |  |  |  |  |
|  | *Estimate* | *95% CI* |  | *Estimate* | *95% CI* |  | *Estimate* | *95% CI* |  |
| ICC | .281 | [.233, .302] |  | .200 | [.165, .219] |  | .263 | [.214, .295] |  |

*Note*. Each independent analysis is weighted by the square root of the sample size relative to the total individuals across all analyses. Confidence intervals (CI) were calculated using 1,000 bootstrapped resamples.

a Variance component could not be estimated.

b By standardizing within each analysis individually systematic mean differences (i.e. variability) across the set of different analyses for the set of criteria are explicitly removed.