# Meta-analysis of second generation competency to stand trial assessment measures: Preliminary findings

Grant A. Blake[a,*], James R.P. Ogloff[a], Won Sun Chen[b]

[a] Centre for Forensic Behavioural Science, Swinburne University of Technology and Forensicare, Melbourne, Australia
[b] Department of Statistics, Data Science and Epidemiology, Swinburne University, Australia

ABSTRACT

Background: Since the 1990s, standardized competency to stand trial assessment measures have been developed to improve the reliability and validity of expert assessment and opinion in legal settings. This study aimed to meta-analyze the psychometric properties of the four 'second generation' assessment tools: the MacCAT-CA, MacCAT-FP, ECST-R and CAST*MR.
Method: A systematic search of published and unpublished studies was undertaken on PsycNet, PubMed, Scopus, Web of Science and Science Direct up until February 2018. Test manuals were sourced and attempts at finding grey literature included contacting study authors and test developers.
Results: Eleven studies were included in the final analysis. The MacCAT-CA, ECST-R and CAST*MR demonstrated acceptable subscale internal consistencies and interrater reliability. The MacCAT-CA and CAST*MR discriminated between competent and incompetent defendants with large effects. The MacCAT-FP had insufficient reliability data and poor discrimination. There was insufficient data to meta-analyze the factor structure, sensitivity, specificity, or effect sizes between types of clinical presentation on any tool.
Conclusions: Very few studies have investigated the psychometric properties of the selected tools. Many studies were excluded for not providing original data or simply re-reporting the psychometric properties stated in test manuals. Whilst the MacCAT-CA and ECST-R have promising psychometric properties, ongoing construct validation is required.

## 1. Introduction

Competency to stand trial refers to a defendant's mental capacity to participate in their adjudication.[1] Although the legal test for incompetence varies across jurisdictions, defendants are generally expected to possess a basic understanding of the charges and legal process, be able to communicate with their legal representatives, and make rational decisions. It is presumed unfair to try the defendant in the usual manner if they lack any one of these abilities.

The English common law construct has evolved into the highest frequency forensic mental health assessment in the United States (Pirelli & Zapf, 2008). This evolution has been aided by the development of assessment tools that improve the reliability, validity and accountability of expert mental health opinions. The earliest assessment tools comprised checklists of basic legal abilities (Robey, 1965), sentence completion tasks (Lipsitt, Lelos, & McGarry, 1973), and semi-structured interview (Roesch, Webster, & Eaves, 1984).

Although accountability and reliability were improved through standardization, empirical research identified significant limitations to the construct validity of these tools. For example, the Georgia Court Competency Test (GCCT), Fitness Interview Test - Revised (FIT-R) and Competency Screening Test (CST) have all demonstrated unstable factor structures (Bagby, Nicholson, Rogers, & Nussbaum, 1992; Rogers, Ustad, Sewell, & Reinhardt, 1996). The GCCT-Mississippi State Hospital revision demonstrated high misclassification rates (Ustad, Rogers, Sewell, & Guarnaccia, 1996), and the CST has failed to discriminate between competent and incompetent defendants (Menzies, Webster, Roesch, Jensen, & Eaves, 1984). Nonetheless, these instruments have stronger correlations with competency than traditional assessments, such as personality and cognitive ability tests (Nicholson & Kugler, 1991; Pirelli, Gottdiener, & Zapf, 2011). The first competency measures were therefore relevant to addressing the legal question but

---

* Corresponding author.
  E-mail address: gablake@swin.edu.au (G.A. Blake).
[1] Common law jurisdictions outside of the United States refer to Competence to Stand Trial as Fitness to Stand Trial (e.g., Australia, Canada, Ireland, New Zealand, Singapore, United Kingdom), although the concepts are very similar. As with CST, Fitness to Stand Trial criteria vary somewhat across jurisdictions.

required ongoing refinement.

## 1.1. Second-generation competency measures

To improve standardization and address threats to validity, several *second-generation* competency assessment tools have been developed. The notion of second-generation tools was coined by Grisso (1991) who called for a new suite of competency interviews to overcome the limitations of those previously developed. He explained that while first generation tools helped explicate the competency construct, second generation tools were required to improve how that construct was assessed. For example, Robey's (1965) competency checklist outlined the expected functional capacities of a competent defendant. The second-generation tools standardize how that ability should be assessed, evaluate that ability in relation to the legal criteria, and provide normative data for comparative analysis and interpretation. Similarly, the first-generation measures began standardizing how information should be collected, whilst the second-generation interviews standardized how that information should be scored and interpreted.

The first of the second-generation tools was the Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST*MR; Everington & Luckasson, 1992). This assessment measure has received little attention in the literature (Pirelli et al., 2011) and its infrequent utilisation is reflected in studies evaluating forensic mental health test usage. To date, the CAST*MR has never been rated as a frequently used assessment tool despite competency to stand trial being the most common forensic mental health assessment (Hill & Demetrioff, 2018; Neal & Grisso, 2014). In an earlier study, Lally (2003) found that none of their surveyed clinicians offered an opinion on the CAST*MR. One might presume that this is due to the competency construct evolving as mostly associated with psychosis in US case law.

Several years later, the MacArthur Competence Assessment Tool – Criminal Adjudication (MacCAT-CA; Poythress et al., 1999) was released. The MacCAT-CA was the product of a large scale research project that developed a lengthy competency measure, which was later refined into the MacCAT-CA. The research team comprised industry experts with extensive knowledge of the Dusky criteria, and the validation research involved large samples of prison inmates, forensic mental health inpatients, and forensic mental health patients found incompetent to stand trial (Otto et al., 1998). The MacCAT-Fitness to Plead (MacCAT-FP) was released soon after for use in the United Kingdom. Whilst the MacCAT-FP was subjected to psychometric evaluation, the samples used were much smaller than its predecessor.

Lastly, the Evaluation of Competency to Stand Trial – Revised (ECST-R; Rogers, Tillbrook, & Sewell, 2004) was the product of a doctoral dissertation by the second author. The ECST-R departed from the former measures in several practical and theoretical ways. First, items were developed in consultation with legal experts to ensure they addressed the Dusky criteria. Second, the scoring procedure is the reverse of the former tools. As all defendants are presumed to be competent, the ESCT-R is used to evaluate the defendant's incompetence rather than inferring incompetence from a low competency score. Higher scores on the ECST-R indicate greater likelihood of incompetency. Third, the ECST-R employs semi-structured and structured interview techniques. The latter is used to identify feigned incompetence and malingered symptoms directly related to the defendant's own trial. The ECST-R is the only second-generation competency measure to evaluate malingering and feigning specific the defendant's own trial.

## 1.2. Psychometric meta-analysis

The literature has identified these second generation tools to be of great importance in improving forensic mental health assessments. Several narrative reviews have discussed the strengths and weaknesses of competency measures, generally concluding that second-generation

tools are more comprehensive in their scope and utility (Fogel, Schiffman, Mumley, Tillbrook, & Grisso, 2013). However, there appears to have been no efforts to meta-analyze the psychometric evidence of these tools. This is problematic given that tests require ongoing evaluation to delineate construct validity (Strauss & Smith, 2009). Further, psychometric meta-analysis is recommended when an assessment tool is evidently valuable to a particular field (Hale, Crocetti, Raaijmakers, & Meeus, 2011), much like the newer competency measures have been said to improve evaluations.

There are substantial benefits to meta-analysis compared with other types of review. These include enhancing statements of external validity, describing the reliability of research findings, and synthesising complex information from multiple studies into a single statistic that represents the strength of a relationship between two variables (Gliner, Morgan, & Harmon, 2003). Meta-analysts often source grey literature and non-significant results, thus attempting to overcome publication bias and improve statements of external validity. Psychometric meta-analysis achieves the same conclusory outcome of other types of meta-analysis but is a relatively new analytical technique that has been scripted into evidence-based code (e.g., Metacor package in R). The primary difference between psychometric meta-analysis and other types of meta-analysis is the capacity to analyze correlational statistics, such as Cronbach's alpha.

The superiority of psychometric meta-analysis to narrative reviews also aligns with the *Daubert v. Merrel Dow Pharmaceuticals, Inc.* (1993) evidence admissibility standard. Error rates, such as false positives and false negatives, should be known to evaluators so the court may critically appraise an expert's findings from a given assessment procedure. Problematically, several studies have found that the Courts almost always agree with expert opinion on competency (Crocker, Favreau, & Caulet, 2002; Gowensmith, Murrie, & Boccaccini, 2012) and that legal and mental health professionals often misunderstand one another (Rogers, Blackwood, Farnham, Pickup, & Watts, 2009). Thus, the accretion of psychometric evidence is necessary to inform evaluators about the strengths and limitations of these tools, which may then be succinctly communicated to the courts for consideration.

## 1.3. Current study

The aim of this research was to meta-analyze the psychometric properties of second-generation competency measures for adult defendants. Specifically, we aimed to answer the following questions for the MacCAT-CA, MacCAT-FP, ECST-R and CAST*MR: (1) What are the subscale internal consistencies for each tool; (2) What is the interrater reliability for each tool?; (3) What is the test-retest reliability for each tool?; (4) What is the effect size between competent and incompetent defendants for each tool?; (5) What is the diagnostic sensitivity of each tool?; (6) What is the diagnostic specificity of each tool?, and (7) What is the best fitting factor solution for each tool?

## 2. Method

This review was registered with the PROSPERO International prospective register of systematic reviews (CRD42017065521) and utilised the Preferred Reporting Items for Systematic Reviews and Meta-Analyses method (PRISMA; Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009).

## 2.1. Instruments

### 2.1.1. The MacArthur competence assessment tool – criminal adjudication (MacCAT-CA)

The MacCAT-CA (Poythress et al., 1999) is a 22-item semi-structured interview comprised of three subscales, namely Understanding (8-items), Reasoning (8-items) and Appreciation (6-items). The Understanding and Reasoning items are answered in response to a brief
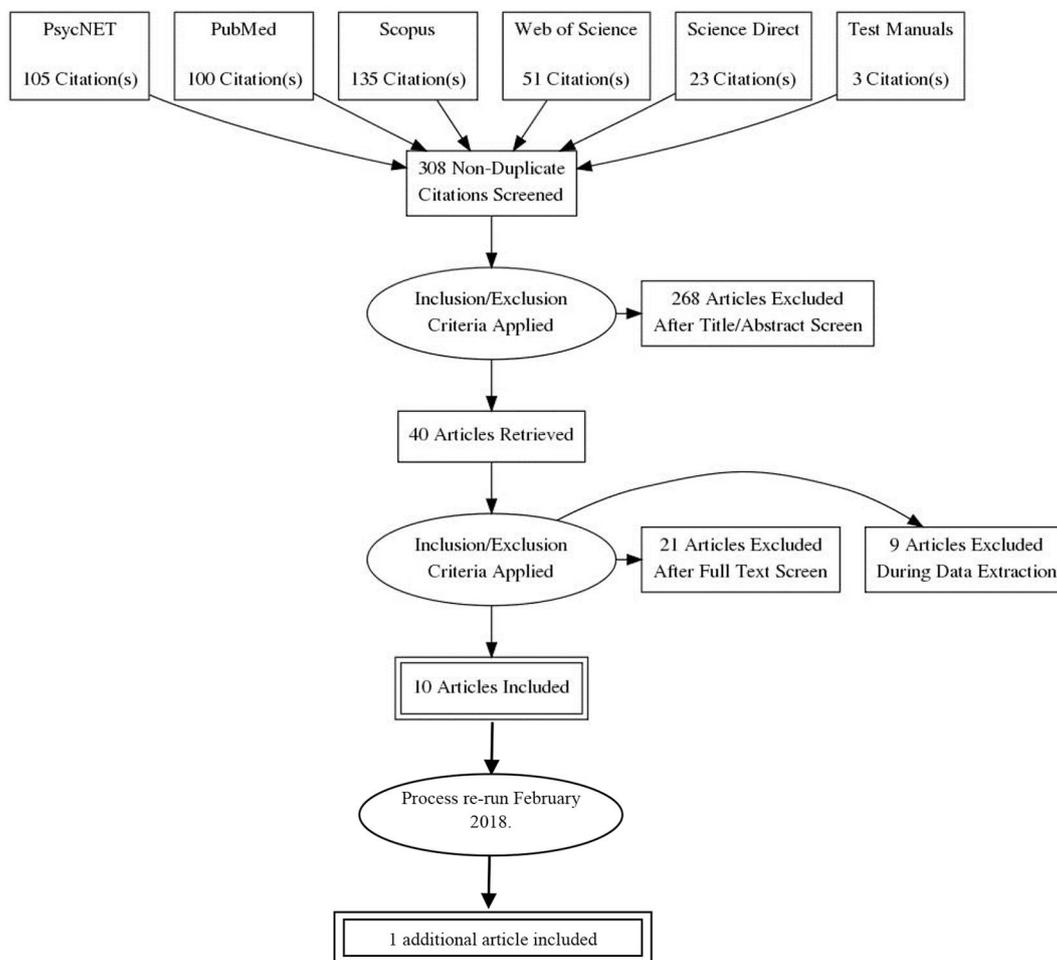
**Fig. 1.** Search strategy results.

**Table 1**
JBI Critical appraisal checklist for diagnostic test accuracy studies.

| | | | | Item | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| First author | Year | Tool | N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Akinkunmi | 2002 | MacCAT-FP | 105 | Y | Y | N | Y | N/A | – | Y | Y | N | Y |
| Everington | 1995 | CAST-MR | 35 | U | N | N | N | Y | – | Y | U | N | Y |
| Everington | 1990 | CAST-MR | 93 | U | U | Y | U | Y | – | Y | Y | Y | Y |
| Jacobs | 2008 | MacCAT-CA | 177[a] | Y | N | Y | Y | N/A | – | Y | U | Y | Y |
| Otto | 1998 | MacCAT-CA | 729 | Y | N | N | N | Y | – | Y | Y | N | Y |
| Redding | 1997 | MacCAT-CA | 29 | Y | Y | N | N/A | Y | – | N/A | N/A | N/A | Y |
| Rogers | 2003 | ECST-R | 411 | – | – | – | – | – | – | – | – | – | – |
| Rutledge | 2008 | MacCAT-FP | 102 | Y | Y | Y | U | N/A | – | U | U | U | Y |
| Ryba | 2011 | MacCAT-CA | 77[a] | Y | U | Y | U | Y | – | Y | Y | Y | Y |
| Vitacco | 2009 | ECST-R | 100 | U | N | Y | Y | Y | – | Y | U | Y | N |
| Wood | 2017 | MacCAT-CA | 103 | Y | N | U | N | Y | – | Y | N | Y | Y |

Note. [a] = 77 of the participants from Jacobs are the same as those from Ryba 2011, Y = Yes, N = No, U = Unclear, N/A = Not applicable, 1 = Random or consecutive sampling used, 2 = Case-control design avoided, 3 Avoided exclusions, 4 = Index test blind to reference test results, 5 = Used a pre-specified threshold, 6 = Reference test likely to classify condition, 7 = Reference test blind to index test, 8 = Appropriate interval between index and reference tests, 9 = Same reference test applied to all patients, 10 = All patients analyzed.

vignette that is read aloud to the defendant. The MacCAT-CA was published in 1999 and is the product of two large research projects and refinement of its predecessor, the MacArthur Structured Assessment of Competence-Criminal Defendants (MacSAC-CD; Hoge, Bonnie, Poythress, & Monahan, 1990).

The MacCAT-CA was designed to address Bonnie's (1992) legal competence theory for defendants aged 18-years and above, and

reflects the Dusky criteria. It takes between 25 and 55-minutes to administer. Normative data come from 729 participants, including competent and incompetent defendants and general inmates. Defendants with intellectual disability, IQs < 60, dementia and other neuro-cognitive disorders, and English as a second language were excluded from the normative research. A reading level is not reported for the MacCAT-CA.

**Table 2**
Subscale Reliabilities, and Total Score and Subscale Mean Differences between Competent and Incompetent Defendants.

| Measure | Subscale | Internal Consistency | Interrater reliability | Mean inter-item correlation | Competent $M$ | Incompetent $M$ | Effect Size |
|---|---|---|---|---|---|---|---|
| MacCAT-CA | Understanding | .84 | .93 | – | 12.22 | 8.18 | 0.93 |
| | Appreciation | .87 | .82 | – | 1.68 | 6.30 | 1.16 |
| | Reasoning | .81 | .93 | – | 12.50 | 8.21 | 1.01 |
| MacCAT-FP | Understanding | – | – | – | – | – | −0.18 |
| | Appreciation | – | – | – | – | – | 0.05 |
| | Reasoning | – | – | – | – | – | −0.18 |
| ECST-R | FAC | .84 | .96 | .51 | – | – | – |
| | RAC | .87 | .95 | .49 | – | – | – |
| | CWC | .88 | .97 | .55 | – | – | – |
| | Rational | .93 | .96 | .51 | – | – | – |
| CAST*MR | VLC | .88 | – | – | – | – | – |
| | AAD | .75 | – | – | – | – | – |
| | RFE | .83 | – | – | – | – | – |
| | Total | – | .91 | – | 35.43 | 25.96 | 2.23 |

- = Insufficient data for meta-analysis. FAC = Factual understanding of the courtroom proceedings, RAC = Rational understanding of the courtroom proceedings, CWC = Consult with counsel, VLC = Vocabulary and concepts, AAD = Ability to assist in own defence, RFE = Relation of factual events.

**Table 3**
Correlations between MacCAT-CA and BPRS Subscales.

| Subscale | Total BPRS | Psychoticism | Depressed | Hostility | Withdrawal |
|---|---|---|---|---|---|
| Understanding | −.23[*] | −.41[*] | 0.05 | −.13 | −.29[*] |
| Appreciation | −.36[*] | −.51[*] | 0.03 | −.28[*] | −.26[*] |
| Reasoning | −.35[*] | −.49[*] | 0.03 | −.24[*] | −.26[*] |

[*] = $p < .001$.

#### 2.1.2. The MacArthur competence assessment tool – fitness to plead (MacCAT-FP)

The MacCAT-FP is a British adaption of the MacCAT-CA. There is no test manual for the MacCAT-FP so limited information is available on how it differs from the MacCAT-CA. Akinkunmi (2002) reported two item amendments: Reference to juries in sentencing is omitted and offence designations were adjusted to reflect appropriate terminologies.

Like the MacCAT-CA, the MacCAT-FP has 22-items that comprise 3 subscales of the same name. Similarly, the first 16-items are answered in response to the same MacCAT-CA vignette that is real aloud to the defendant. Administration time and reading level is unknown. There is no normative dataset for the MacCAT-FP. The MacCAT-FP was analyzed separate from the MacCAT-CA, as the legal test by which clinicians form an ultimate opinion in England is different to the United States.

#### 2.1.3. Evaluation of competency to stand trial (ECST-R)

The ECST-R (Rogers et al., 2004) is an 18-item semi-structured interview comprising three subscales named Factual Understanding of the Courtroom Proceedings (FAC, 6-items), Rational Understanding of the Courtroom Proceedings (RAC, 6-items) and Consult with Counsel (CWC, 6-items). Half of the items have optional follow-up probes that standardize further investigation of the defendant's knowledge and thought processes. Scoring procedures prompt evaluators to rate the presence and impact of psychotic symptoms on the defendant's competency.

The ECST-R evaluates the defendant's competency in relation to their own charges and proceedings. The ECST-R was published in 2004 and normative data come from 356 offenders who passed malingering and feigning measures. Offenders with IQs < 60 were excluded from normative research. The ECST-R was designed to address the Dusky v. United States (1960) competency standards and to satisfy the *Daubert v. Merrel Dow Pharmaceuticals, Inc.* (1993) evidence admissibility standards. It can be administered to adults aged 18-years and above, and takes between 25 and 45-minutes to complete. It has < 8-words on average per item but a reading level is not reported. The ECST-R is the only second generation competency measure with an inbuilt symptom malingering and competency feigning screener. This response style

measure has been the subject of numerous studies that have concluded it is fit for purpose (Fogel et al., 2013).

#### 2.1.4. Competence assessment for standing trial for defendants with mental retardation (CAST*MR)

The CAST*MR (Everington & Luckasson, 1992), previously abbreviated to the CAST-MR, is a standardized competency interview for defendants with intellectual disability in the United States. The CAST*MR has 50-items comprising three subscales named Vocabulary and Concepts (25-items, multiple choice response), Ability to Assist in Own Defense (15-items, multiple choice response), and Relation of Factual Events (10-items, open ended questions). The CAST*MR was published in 1992, takes between 30 and 45-minutes to administer, and was written at a fourth grade level. It can be administered to defendants with a moderate or mild intellectual disability aged 18-years and above.

#### 2.2. Literature search

Electronic searches were performed on 20th March 2017. The databases accessed were: PsycNet (including PsycInfo, PsycBooks and PsycArticles), PubMed, Scopus, Web of Science and Science Direct. Reference lists of the included articles and competency research reviews (Cooper & Grisso, 1997; Fogel et al., 2013; Grisso, 1992; Mumley, Tillbrook, & Grisso, 2003) were hand-searched by the first author. No date range was specified and published and unpublished studies were included.

The search terms were developed to be consistent with other psychometric meta-analyses and competency research (e.g., Pirelli et al., 2011; Scaini, Battaglia, Beidel, & Ogliari, 2011). Scoping searches were conducted to refine the search terms to reduce the number of irrelevant citations whilst maintaining accurate identification of possibly relevant citations. The final search term was ("cast-mr" OR "cast*mr" OR "Competence Assessment for Standing Trial for Defendants with Mental Retardation") OR ("ECST-R" OR "evaluation of competency to stand trial") OR ("Maccat-ca" OR "maccat-fp" OR "MacArthur Competence Assessment Tool") AND ("stand trial") AND NOT ("juvenile" OR "adolescent" OR "adolescence"). The searches were re-run on 19th February 2018 to identify recently published articles.

User manuals were sourced for the MacCAT-CA, ECST-R and CAST*MR. Study authors and test publishers were emailed to request unpublished data, missing data, and other grey literature.

#### 2.3. Study selection

The first author conducted the screening process then applied the inclusion and exclusion criteria. Articles were included if they used at
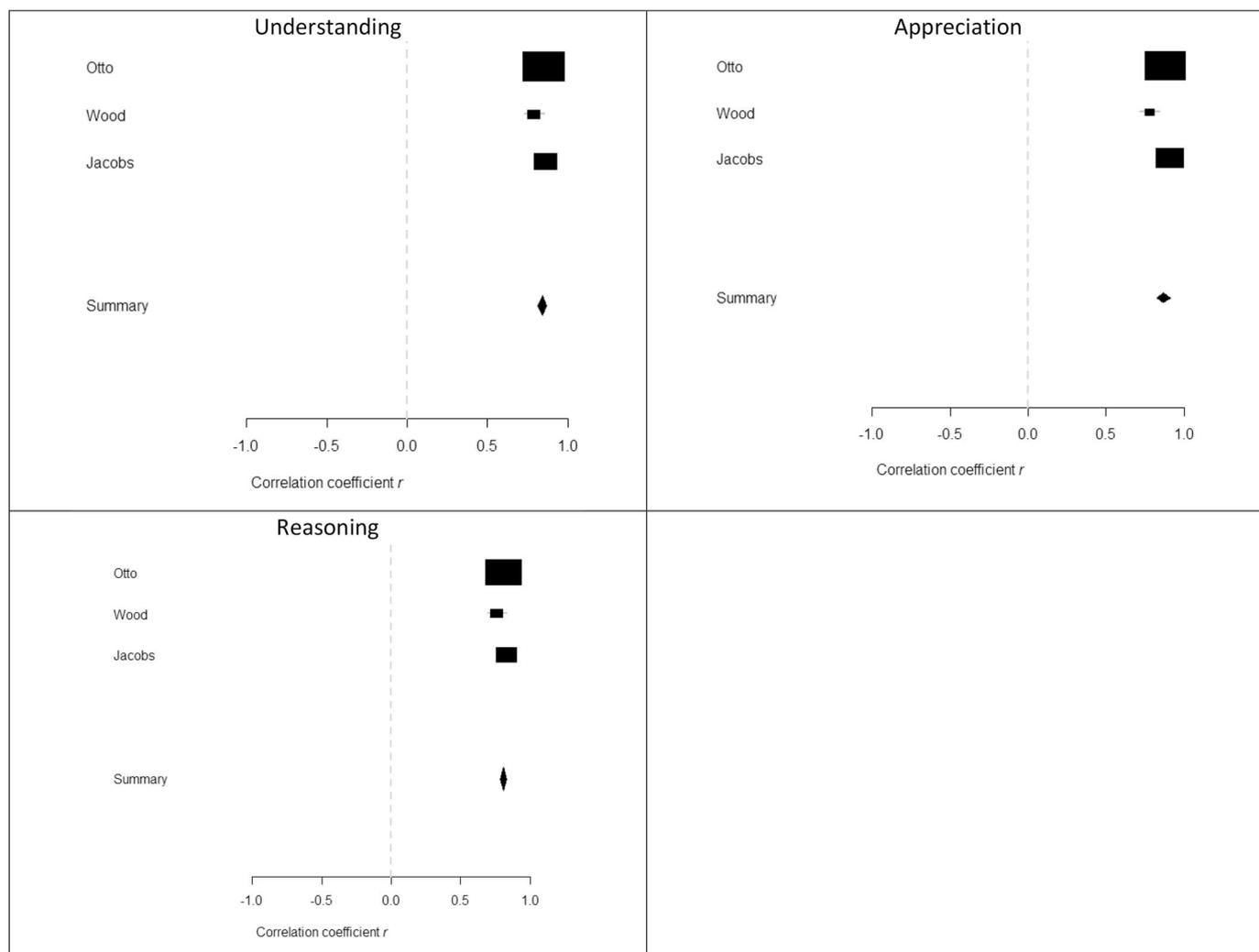
**Fig. 2.** MaCCAT-CA subscale internal consistencies.

least one of the four assessment tools, compared competent and incompetent defendants, provided new data, and reported either means and standard deviations, reliability estimates, predictive power estimates, or factor solutions. Studies that re-analyzed previously published data or evaluated malingering only were excluded.

### 2.4. Quality analysis

Study quality was initially evaluated with the competency research checklist developed by White, Meares, and Batchelor (2014); however, a high number of discrepancies occurred between the first author and research assistant. Therefore, second quality analysis was undertaken utilising the Critical Appraisal Checklist for Diagnostic Test Accuracy Studies by the Joanna Briggs Institute (Campbell et al., 2015; Whiting et al., 2011). The checklist is comprised of ten items that evaluate risk of sampling bias, index test and reference test bias, temporal bias, and analytical bias. Item 6 was omitted from the present study as there is no gold-standard diagnostic reference test for competency (Gowensmith et al., 2012). There was perfect coding agreement between the first author and research assistant ($\kappa = 1.0$).

### 2.5. Data extraction

Data were extracted from articles by the first author and research assistant with perfect reliability ($\kappa = 1.0$). Only the first author extracted data from the test manuals. The coding guide was developed in

an iterative manner to ensure inclusion of relevant information not identified a-priori. The extracted information included study methods, subscale Cronbach's alphas and/or Kuder-Richardson statistics, mean inter-item correlations, inter-rater reliability estimates and parameters (i.e., number of raters and number of co-rated protocols), test-retest reliability estimates, rates of incompetency for the total sample and by diagnosis type (e.g., psychotic, intellectual disability), and the single best supported factor solution within a study with accompanying factor loadings, errors and fit indices (e.g., NNFI, CFI, TLI, SRMR).

Means and standard deviations for competent and incompetent groups were also coded, as well as means and standard deviations for diagnostic groups (e.g., psychotic and non-psychotic, intellectual disability and no intellectual disability) and by gender. Lastly, several of the MacCAT-CA studies investigated concurrent validity by correlating competency scores with scores from the Brief Psychiatric Rating Scale. The simple Pearson correlations from these studies were recorded.

### 2.6. Analytic plan

The data were analyzed in R (R Core Team, 2013). A random effects model was used to adjust for differences in study characteristics, per standard meta-analytic procedures (Viechtbauer, 2010). Specifically, the metacor.DSL function was utilised to implement the DerSimonian-Laird (DSL) random-effect meta-analytical approach with correlation coefficients as effect sizes (Schulze, 2004). Additional effect sizes were computed from raw data using the effect_sizes function in the esc
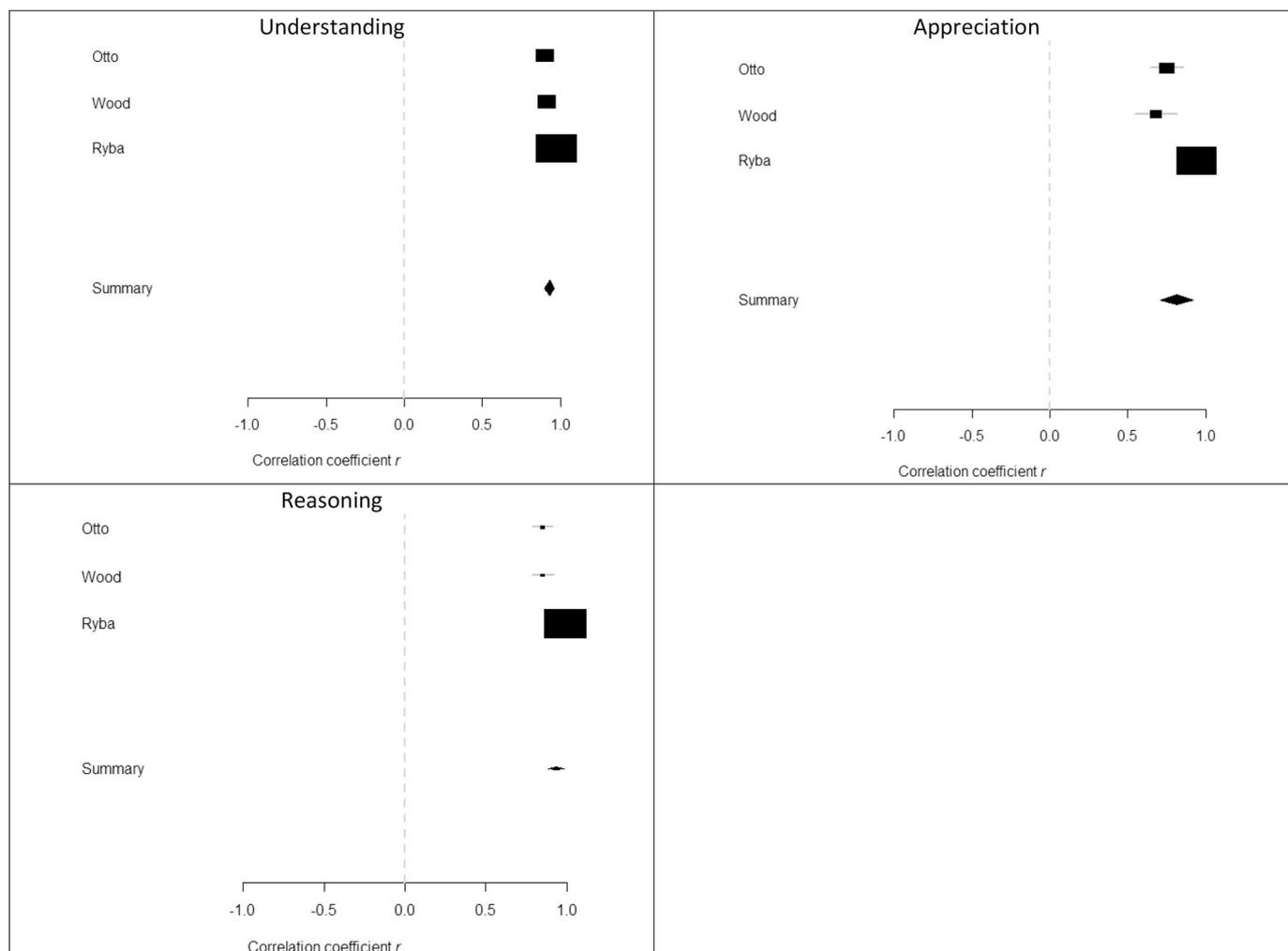
**Fig. 3.** MacCAT-CA subscale inter-rater reliabilities.

package (Ludecke, 2018). Forest plots were computed to illustrate the results and precision of individual study estimates (Guddat, Grouven, Bender, & Skipka, 2012; Higgins, Thompson, & Spiegelhalter, 2009).

## 3. Results

### 3.1. Included studies

The search strategy resulted in 417 citations, 308 of which were non-duplicates. Eleven studies met the criteria for inclusion in the analysis (see Fig. 1). Each assessment tool had two reports except the MacCAT-CA which had four (Table 1). The three individual studies that comprise the psychometric properties section of the ECST-R manual were unavailable, so the manual was coded as a single report. The MacCAT-CA and CAST*MR manuals reported the same results as those in the identified articles, so the data were not coded again. The Wood, Anderson, and Glassmire (2017) study was captured when the search was re-run in 2018. Individual sample sizes ranged from 35 participants to 729 participants. Seventy-seven participants from Jacobs, Ryba, and Zapf (2008) are the same as those from Ryba and Zapf (2011). The former reported Cronbach's alpha whereas the latter did not. Both studies were included but only unique results were included in the meta-analysis.

Nine studies were excluded at the data extraction phase due to re-using or re-analysing previous results (Marcus, Poythress, Edens, & Lilienfeld, 2010; Zapf, Skeem, & Golding, 2005); not reporting any new

psychometric properties (Edens, Poythress, Nicholson, & Otto, 1999; Mueller & Wylie, 2007; Viljoen, Zapf, & Roesch, 2004; Zapf & Roesch, 2001), or the psychometric results included scores from coached feigners (Norton & Ryba, 2010). One study was excluded due to very poor reporting (Jurecska, Peterson, & Millkey, 2012).

### 3.2. Quality analysis

Results from the quality analysis are in Table 1. Overall, almost all studies used a pre-determined or statistically derived cut-score, test scores were not known to independent second opinions (e.g., psychiatrist, court outcome), and all participants were included in the analyses. The highest quality study was by Ryba and Zapf (2011) which achieved 7 out of 9 positive checks. The lowest quality study was by Everington and Dunn (1995), which achieved 3 out of 9 positive checks.

Quality of the ECST-R manual could not be coded, as the results presented are an amalgamation of results from several unpublished doctoral studies. The thesis containing those studies could not be located via ProQuest Central, EBSCO Open Dissertation, nor personal communication. Appreciably, the manual does not address the methodological intricacies typically reported in peer-reviewed literature, and the variety of methods used in each study cannot be summarised across a single analysis. As such, a quality analysis of the ECST-R manual was not undertaken as it would not have been a valid representation of the research.
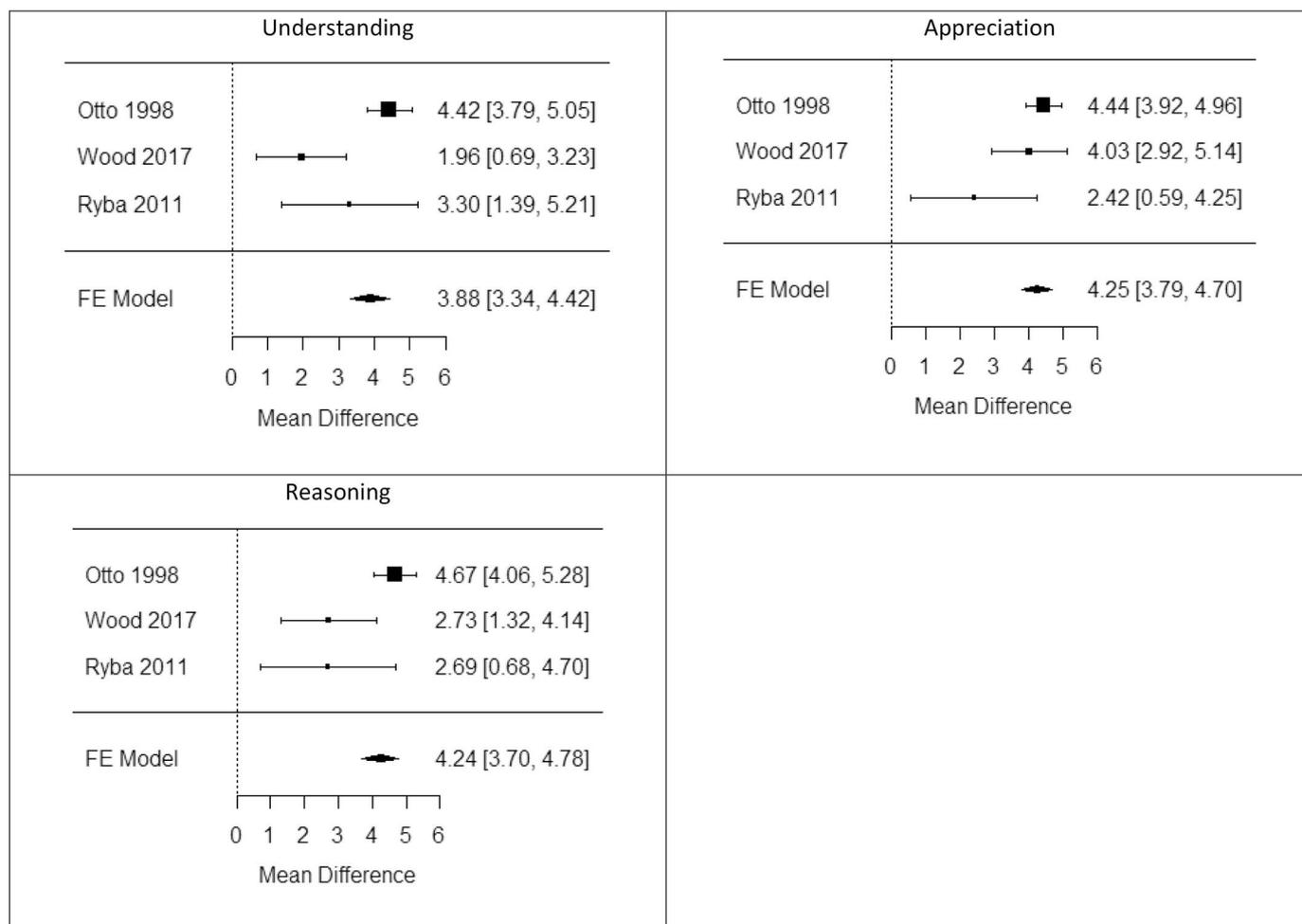
**Fig. 4.** MacCAT-CA subscale mean differences.

### 3.3. Psychometric properties

Results for the psychometric meta-analysis are presented in Tables 2 and 3 and Figs. 2–9. Accepted standards were used to interpret internal consistency (Cronbach, 1951), interrater reliability (McHugh, 2012), mean inter-item correlations (Clark & Watson, 1995), and effect sizes (Cohen, 1988; Sawilowsky, 2009).

#### 3.3.1. MacCAT-CA

The four MacCAT-CA studies had a total sample of 1038 participants. The MacCAT-CA achieved good internal consistency and almost perfect interrater reliability for each subscale. The mean difference effect size between competent and incompetent defendants was consistently large, with evidence of slightly greater effect on the Appreciation subscale than Understanding and Reasoning. Correlations between the Brief Psychiatric Rating Scale (BPRS) and MacCAT-CA (Table 3) subscales found a consistently inverse, moderate strength relationship between competency and psychoticism, and a non-significant, negligible relationship with depression. Emotional withdrawal had a weak relationship with each subscale, whereas hostility was weakly correlated with only two of the MacCAT-CA subscales.

#### 3.3.2. MacCAT-FP

The two MacCAT-FP studies had a total sample of 207 participants. There was insufficient data to meta-analyze any reliability estimate on the MacCAT-FP. Small effects between competent and incompetent defendants were evident on the Understanding and Reasoning subscales, with essentially no difference between competent and incompetent defendants on the Appreciation subscale. The MacCAT-FP was the only tool with sufficient data to meta-analyze an AUC. The MacCAT-FP achieved acceptable predictive validity (AUC = 0.80).

#### 3.3.3. ECST-R

The two ECST-R reports had a total sample of 511 participants. The ECST-R achieved good to excellent internal consistency and almost perfect interrater reliability for each subscale. Mean inter-item correlations per subscale ranged from 0.49 to 0.55. For narrow constructs, scores between 0.40 and 0.50 are considered desirable. Scores above this might indicate item redundancy. There was insufficient data to compare competent and incompetent defendant scores.

#### 3.3.4. CAST*MR

The two CAST*MR studies had a total sample of 128 participants. The CAST*MR achieved acceptable to good internal consistency and almost perfect interrater reliability as a whole scale. There was a very large effect of competency on total scores.

#### 3.3.5. Untested questions

Several of the research questions could not be investigated due to insufficient data and/or reporting practices. There was insufficient test-retest reliability data, gender comparisons, diagnostic comparisons (e.g., mean psychotic versus mean non-psychotic), factor analytic data (e.g., NNFI), and sensitivity and specificity estimates for any of the measures. Lastly, no factor solutions were meta-analyzed due to a tendency to only report significant factor loadings.
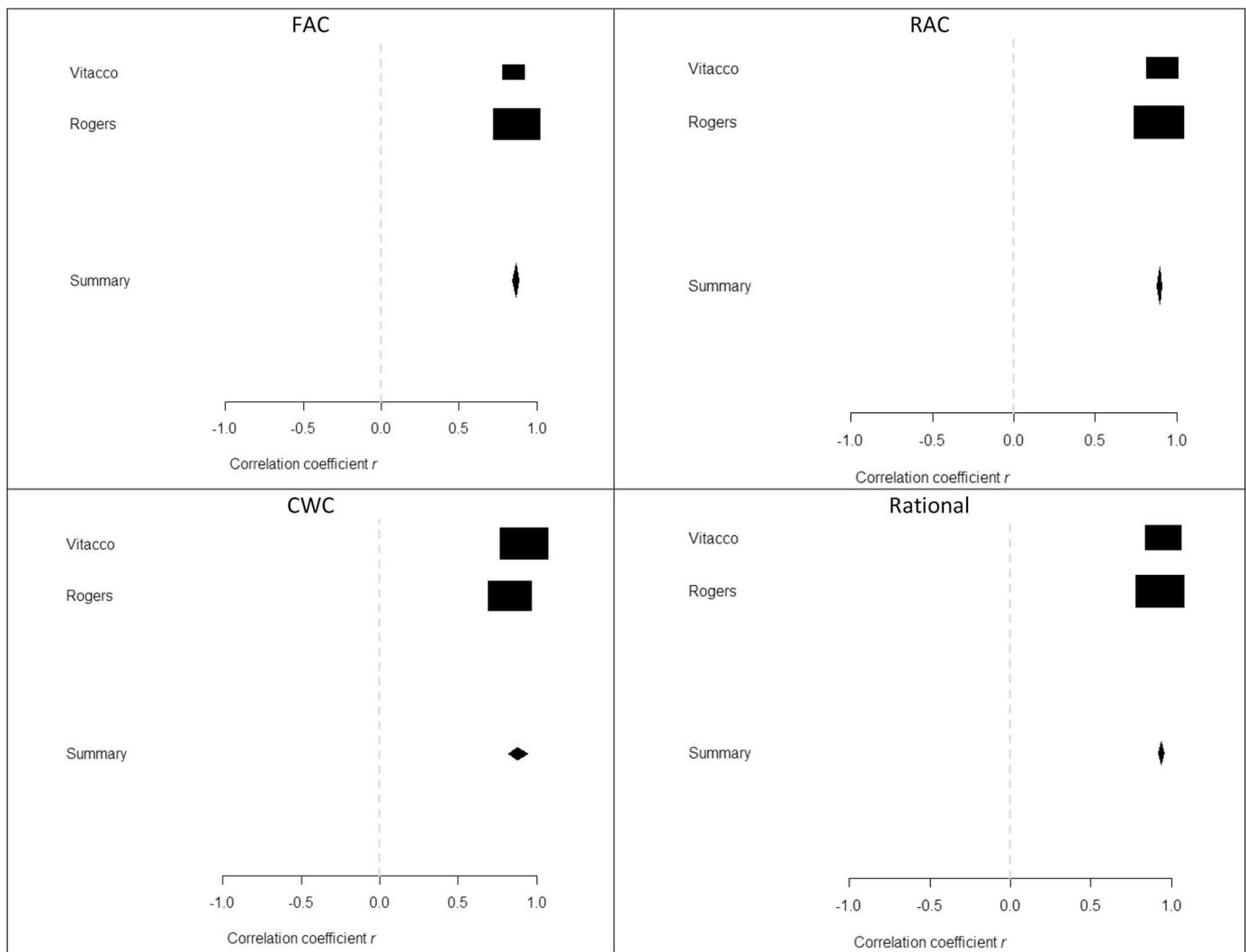
Fig. 5. ECST-R subscale internal consistencies.

## 4. Discussion

This study aimed to meta-analyze the psychometric properties of second-generation competency measures. Overall, the results indicated promising psychometric properties for the MacCAT-CA, ECST-R and CAST*MR but not the MacCAT-FP. Internal consistencies were good across the three measures. Interrater reliabilities spanned across the upper and lower limits of the almost perfect range (McHugh, 2012).

The psychometric results indicated that the ECST-R had slightly stronger reliability than the MacCAT-CA and CAST*MR, although this could be inflated due to potential redundancy within some of the ECST-R subscales. Unfortunately, there were no indicators of potential redundancy for the MacCAT-CA and CAST*MR that would enable consideration of similar issues. Regarding capacity to discriminate between competent and incompetent defendants, the CAST*MR total score outperformed the MacCAT-CA subscales with approximately double the effect size. However, the unavailability of subscale data on the CAST*MR and total score data on the MacCAT-CA means that a clearer comparison cannot be made at this stage. As the CAST*MR is an intellectual disability driven assessment and the MacCAT-CA and ECST-R are psychosis driven instruments, comparative statements should be considered cautiously. The strength of association between psychosis and the MacCAT-CA subscales was the strongest of the correlational analyses, followed by withdrawal symptoms, then hostility on the Appreciation and Reasoning scales only. Depressive symptoms were

unrelated to all subscales.

Perhaps the most influential finding from this research is that very few studies met the inclusion criteria for meta-analysis. Three of the four measures only had two data sources. Whilst this meets the minimum criteria for meta-analysis provided that a random-effects model is utilised (Friede, Röver, Wandel, & Neuenschwander, 2017), it is by no means an extensive analysis in the present circumstances. This causes the results to be preliminary in nature. Several excluded studies suggested the competency measures were 'reliable' and 'valid' by re-reporting the psychometric estimates from earlier research. This is problematic because consumers of the newer research are limited in their ability to critically appraise the newer study's finding. It is an inappropriate assumption that evidence of scale reliability from an earlier study presumes scale reliability under new testing circumstances.

This issue might reflect a misunderstanding of the terms 'reliable' and 'valid' as achievable, binary outcomes (i.e., a measure is either reliable or it is not). As constructs are inherently unobservable and are therefore discerned by the nature of their relationships with other variables (e.g., internal consistency, convergent and divergent validity), delineating construct validity is an ongoing process (Strauss & Smith, 2009). Furthermore, every study has inherent limitations with the added possibility of Type I and II errors. The language of a 'reliable' and 'valid' competency measure may therefore be misleading. Assessment tools should be described as having research that supports the notion of
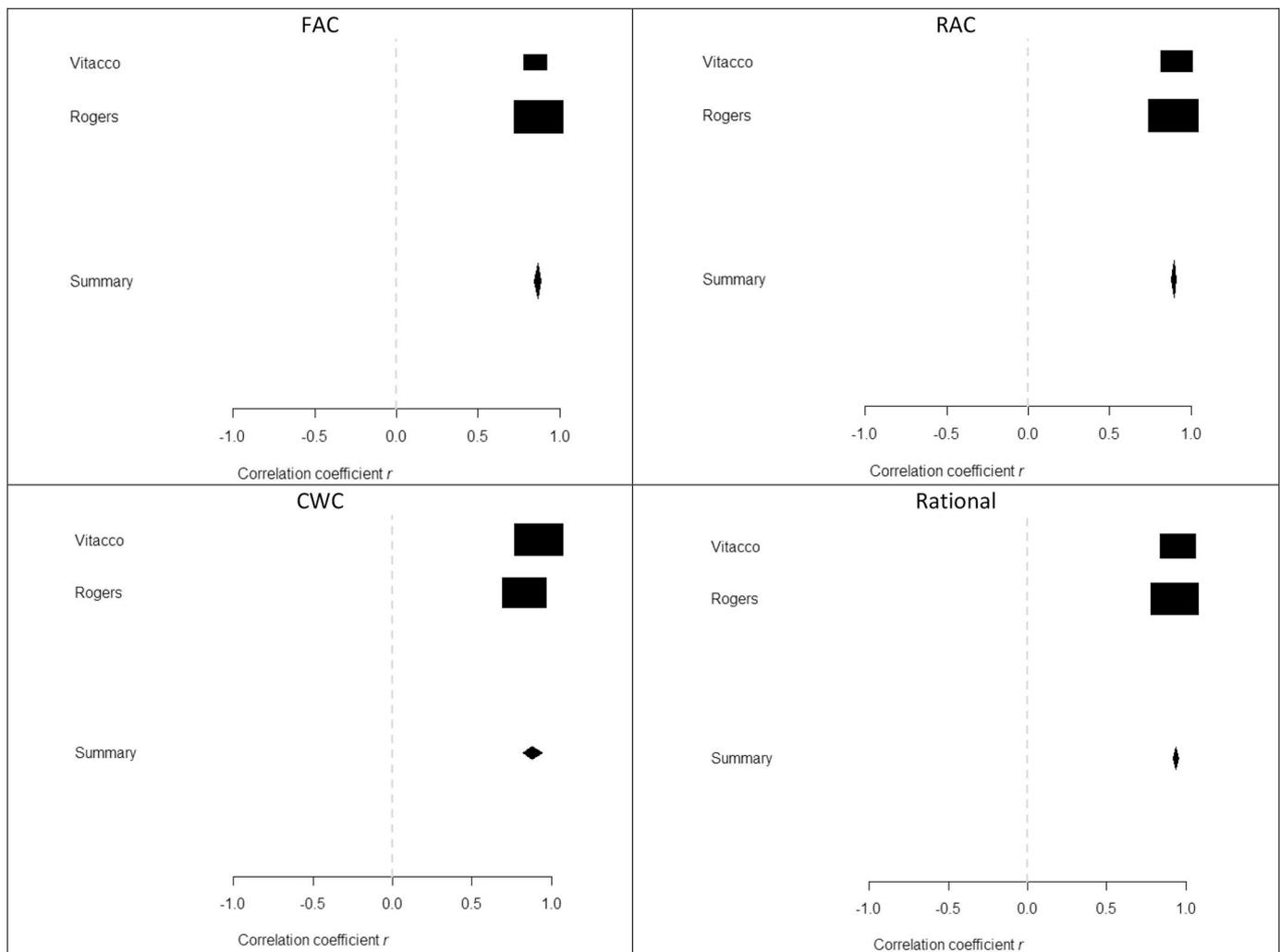
**Fig. 6.** ECST-R subscale inter-rater reliabilities.

their reliability and validity under certain methodological procedures.

Moreover, failing to report reliability estimates deprives future researchers of the opportunity to undertake meta-analysis, or can cause researchers to impute or exclude primary effects from their own analyses (Plonsky & Derrick, 2016). Taken together, it limits the forensic clinician's scope to knowingly provide evidence-based assessments supported by the highest standard of scientific information. This can have drastic consequences for due process in competency cases and may compound the United Nations concerns that competency evaluations often discriminate against persons with a disability (Gooding and O'Mahony, 2016).

In these situations, it requires reminding that the modern construct of competency exists to protect defendants from unfairness, uphold the integrity of the legal system, and ensure reliable verdicts in the public's interest (Grisso, 2014). The ultimate decision regarding a defendant's competency is inextricably influenced by, if not heavily relied upon, the forensic mental health expert's assessment. Continual efforts are therefore required to improve the assessment procedures that inform these opinions as a matter of public interest and human rights. The sparse number of articles that satisfied the inclusion criteria, or even came close to inclusion, emphasises that greater efforts are needed to continue refining and developing jurisdiction specific competency measures. Competency is a global construct and only two countries have second-generation measures.

A surprising finding from this research is the stark difference between the MacCAT-CA and MacCAT-FP effect sizes. Competency status

had a large effect on the MacCAT-CA subscales but a small to negligible effect on the MacCAT-FP subscales. The former is a positive finding whereas the latter raises concern. Whilst there is very little information about the process of modifying the MacCAT-CA to meet UK standards, there is almost no difference between the interviews. Both interviews use the same hypothetical scenario and feature the same number of items on each subscale. There are several possibilities as to why the MacCAT-FP performed poorly in this regard. This includes small sample sizes, random error, and the theoretical framework underpinning these tools.

The MacCAT-CA has been criticized for evaluating Bonnie's (1990, 1992) legal theory of competence rather than the Dusky criteria specifically (Rogers et al., 2004). The MacCAT-FP may have failed to discriminate between competent and incompetent defendants on the basis that test scores are a reflection of American Dusky theory rather than the UK fitness to stand trial criteria. The psychometric assertion that a measure must be deeply grounded in (1) evidence-based theory of a construct, (2) with clear face validity and (3) expert appraised content appears to hold true (Cronbach & Meehl, 1955). The MacCAT-CA was developed to meet these standards and was the most comprehensively evaluated tool with regard to number of publications and total sample size. Conversely, the MacCAT-FP appears to have undergone only minor amendments with consideration of legal process in replacement of expert consensus that it addresses the appropriate construct. Inclusion of the MacCAT-FP in this study, despite the absence of a test manual and normative dataset, was imperative given that it is the only second-
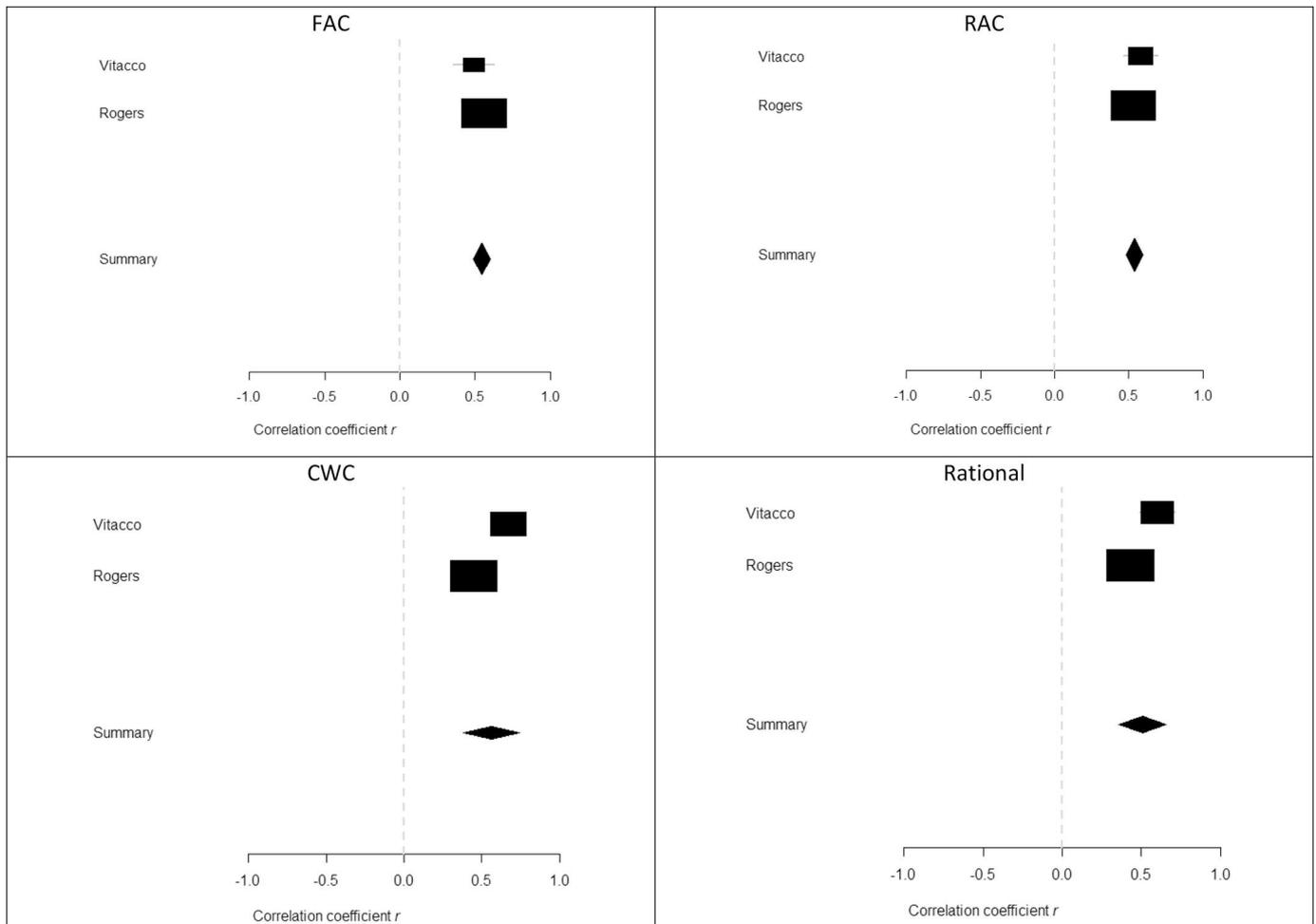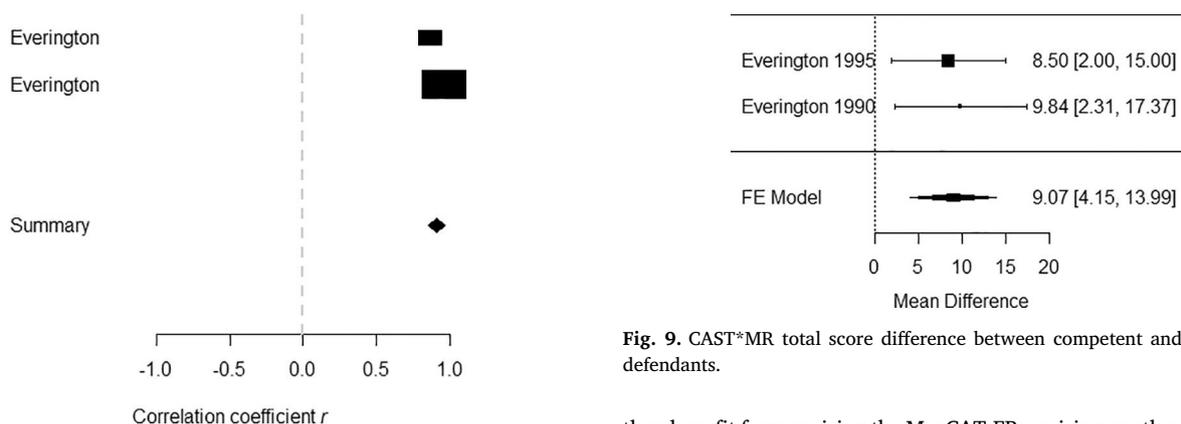
**Fig. 7.** ECST-R subscale mean inter-item correlations.



**Fig. 8.** CAST*MR internal consistency.



**Fig. 9.** CAST*MR total score difference between competent and incompetent defendants.

generation competency measure in England, is based on a measure developed by a reputable group, and met the inclusion criteria.

A small body of literature has criticized the involvement of mental health practitioners in competency evaluations. Szasz (1988) expressed that forensic mental health evaluators should be limited to symptom descriptions only. He complained that mental health clinicians treat competency as a clinical question, have too much influence on legal outcomes, and that competency is ultimately a legal enquiry to be answered by a legal expert. Thus the construct of interest is the legal test and not a theory of the criteria that make up that test. The UK might

then benefit from revising the MacCAT-FP, revising another measure, or developing an entirely new measure to address the Pritchard criteria upon which the legal test is made. Efforts appear to have been made in this regard (Brown et al., 2018), although somewhat surprisingly with the use of a hypothetical scenario again.

Positively, a benefit to the ECST-R items is that they were designed with expert consensus to evaluate the Dusky standard and prompt evaluation of the defendant's abilities as they pertain to their own defense. The test manual also lists numerous legal and mental health experts who assisted in evaluating the test items, thus addressing the psychometric expectation of expert appraised content and face validity. Unfortunately, there was insufficient data to meta-analyze competent and incompetent ECST-R scores, which may have enabled comparison with the other measures' discriminative abilities. Additionally, the

doctoral studies that informed the normative data were unavailable for quality analysis, thus prohibiting an evaluation of the methods used to establish that dataset.

Nonetheless, second generation competency measures appear to provide more accurate evaluations of competency than traditional clinical tools and are increasing in popularity. In a survey (Neal & Grisso, 2014) of forensic clinicians from the US, Canada, Australia, New Zealand and Europe, competency to stand trial was the most common referral question. Among the most commonly used measures were the ECST-R, MacCAT-CA and Juvenile Adjudicative Competence Interview (JACI; Grisso, 2005). Reasons for using assessment tools were predominantly for credibility, evidence-based practice, and method standardization.

From a legal perspective, the Court may have greater trust in an expert's opinion if it was derived from an evidence-measure that can be scrutinized. Unstructured clinical assessments are simultaneously easier and harder to cross-examine. It is easier to discount the expert's opinion based on evidence-admissibility standards and the lack of data supporting their opinion, yet harder for the Court to determine how the defendant was assessed. Although clinicians cannot divulge test item content due to copyright laws, there is a greater sense of accountability and provability when a clinician can describe their procedures in detail and with a statement of being evidence-based.

Structured interviewing, however, is not without limitations. Structured interviews often prohibit the clinician's scope to investigate certain lines of enquiry within the validated assessment protocol. Such is the benefit of semi-structured interviews that standardize the initial lines of enquiry and allow the clinician to continue assessing within their skillset. The second-generation competency interviews employ a semi-structured interview technique, thus providing an accountable, evidence-based method that enables clinician's to employ expert interviewing and formulation skills where necessary.

This study was limited by the small number of included studies for each assessment tool, such as the ECST-R and CAST*MR for which there are only two publications and by the same research groups. We had no success in sourcing grey literature and only one author applied the inclusion and exclusion criteria. Whilst these limitations make the findings preliminary, this study summarised the psychometric evidence behind each measure and makes practical recommendations for how research and clinical practice might be improved.

## 5. Conclusion

This study found that the MacCAT-CA, ECST-R and CAST*MR have promising psychometric properties. Researchers utilising these tools should always report basic psychometric properties to assist in improving clinical assumptions of internal and external validity. Reporting means and standard deviations for demographic and clinical comparisons would also enable the development of a normative dataset for special populations. Further research is required to refine and continue establishing the construct validity of these measures.

## Acknowledgement

## References

Akinkunmi, A. A. (2002). The MacArthur competence assessment tool-fitness to plead: A preliminary evaluation of a research instrument for assessing fitness to plead in England and Wales. *The Journal of the American Academy of Psychiatry and the Law, 30*, 476–482.

Bagby, M. R., Nicholson, R. A., Rogers, R., & Nussbaum, D. (1992). Domains of competency to stand trial: A factor analytic study. *Law and Human Behavior, 16*, 491–507. https://doi.org/10.1007/BF01044620.

Bonnie, R. J. (1990). The competence of criminal defendants with mental retardation to participate in their own defense. *The Journal of Criminal Law and Criminology, 81*, 419–446. https://doi.org/10.2307/1143845.

Bonnie, R. J. (1992). The competence of criminal defendants: A theoretical reformulation. *Behavioral Sciences & the Law, 10*, 291–316. https://doi.org/10.1002/bsl.2370100303.

Brown, P., Stahl, D., Appiah-Kusi, E., Brewer, R., Watts, M., Peay, J., & Blackwood, N. (2018). Fitness to plead: Development and validation of a standardised assessment instrument. *PLoS ONE, 13*, e0194332. https://doi.org/10.1371/journal.pone.0194332.

Campbell, J. M., Klugar, M., Ding, S., Carmody, D. P., Hakonsen, S. J., Jadotte, Y. T., ... Munn, Z. (2015). Diagnostic test accuracy: Methods for systematic review and meta-analysis. *International Journal of Evidence-Based Healthcare, 13*, 154–162. https://doi.org/10.1097/xeb.0000000000000061.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale measurement. *Psychological Assessment, 7*, 309–319. https://doi.org/10.1037//1040-3590.7.3.309.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). (New York, New York: Routledge).

Cooper, D. K., & Grisso, T. (1997). Five year research update (1991-1995): Evaluations for competence to stand trial. *Behavioral Sciences & the Law, 15*, 347–364. https://doi.org/10.1002/(sici)1099-0798(199722/06)15:3<347::aid-bsl270>3.0.co;2-k.

Crocker, A. G., Favreau, O. E., & Caulet, M. (2002). Gender and fitness to stand trial: A 5-year review of remands in Québec. *International Journal of Law and Psychiatry, 25*, 67–84. https://doi.org/10.1016/s0160-2527(01)00089-9.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. https://doi.org/10.1007/bf02310555.

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

Edens, J. F., Poythress, N. G., Nicholson, R. A., & Otto, R. K. (1999). Effects of state organizational structure and forensic examiner training on pretrial competence assessments. *The Journal of Behavioral Health Services and Research, 26*, 140–150. https://doi.org/10.1007/BF02287487.

Everington, C., & Dunn, C. (1995). A second validation study of the competence assessment for standing trial for defendants with mental retardation (CAST-MR). *Criminal Justice and Behavior, 22*, 44–59. https://doi.org/10.1177/0093854895022001004.

Everington, C., & Luckasson, R. (1992). *Competence assessment for standing trial for Defedants with mental retardation: CAST*MR.* Worthington, OH: International Diagnostic Services.

Everington, C. T. (1990). The competence assessment for standing trial for defendants with mental retardation (CAST-MR): A validation study. *Criminal Justice and Behavior, 17*, 147–168. https://doi.org/10.1177/0093854890017002001.

Fogel, M. H., Schiffman, W., Mumley, D. M., Tillbrook, C., & Grisso, T. (2013). Ten year research update (2001 – 2010): Evaluations for competence to stand trial (adjudicative competence). *Behavioral Sciences & the Law, 31*, 165–191. https://doi.org/10.1002/bsl.2051.

Friede, T., Röver, C., Wandel, S., & Neuenschwander, B. (2017). Meta-analysis of two studies in the presence of heterogeneity with applications in rare diseases. *Biometrical Journal. Biometrische Zeitschrift, 59*, 658–671. https://doi.org/10.1002/bimj.201500236.

Gliner, J. A., Morgan, G. A., & Harmon, R. J. (2003). Meta-analysis: Formulation and interpretation. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*, 1376–1379. https://doi.org/10.1097/01.chi.0000085750.71002.01.

Gooding, P., & O'Mahony, C. (2016). Laws on unfitness to stand trial and the UN convention on the rights of persons with disabilities: Comparing reform in England, Wales, Northern Ireland and Australia. *International Journal of Law, Crime and Justice, 44*, 122–145. https://doi.org/10.1016/j.ijlcj.2015.07.002.

Gowensmith, W. N., Murrie, D. C., & Boccaccini, M. T. (2012). Field reliability of competence to stand trial opinions: How often do evaluators agree, and what do judges decide when evaluators disagree? *Law and Human Behavior, 36*, 130–139. https://doi.org/10.1037/h0093958.

Grisso, T. (1991). Clinical assessments for legal decision making: Research recommendations. In S. A. Shah, & B. D. Sales (Eds.). *Law & mental health: Major developments and research needs.* Washington, DC: Government Printing Office.

Grisso, T. (1992). Fiver-year research update (1986 – 1990): Evaluations for competence to stand trial. *Behavioral Sciences & the Law, 10*, 353–369. https://doi.org/10.1002/bsl.2370100306.

Grisso, T. (2005). *Evaluating juveniles' adjudicative competence: A guide for clinical practice.* Sarasota, FL: Professional Resource Press.

Grisso, T. (2014). *Competence to stand trial evaluations: Just the basics.* Sarasota, FL: Professional Resource Press.

Guddat, C., Grouven, U., Bender, R., & Skipka, G. (2012). A note on the graphical presentation of prediction intervals in random-effects meta analyses. *Systematic Reviews, 1*, 34. https://doi.org/10.1186/2046-4053-1-34.

Hale, W. W., Crocetti, E., Raaijmakers, Q. A. W., & Meeus, W. H. J. (2011). A meta-analysis of the cross-cultural psychometric properties of the screen for child anxiety related emotional disorders (SCARED). *Journal of Child Psychology and Psychiatry, 52*, 80–90. https://doi.org/10.1111/j.1469-7610.2010.02285.x.

Higgins, J. P. T., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analyses. *Journal of the Royal Statistical Society, Series A, 172*, 137–159. https://doi.org/10.1111/j.1467-985X.2008.00552.x.

Hill, D., & Demetrioff, S. (2018). Clinical-forensic psychology in Canada: A survey of practitioner characteristics, attitudes, and psychological assessment practices. *Canadian psychology/psychologie canadienne* Advance online publication https://doi.org/10.1037/cap0000152.

Hoge, S. K., Bonnie, R. J., Poythress, N., & Monahan, J. (1990). *The MacArthur competence*

assessment tool - criminal adjudication. Odessa, FL: Psychological Assessment Resources.

Jacobs, M. S., Ryba, N. L., & Zapf, P. A. (2008). Competence-related abilities and psychiatric symptoms: An analysis of the underlying structure and correlates of the MacCAT-CA and the BPRS. *Law and Human Behavior, 32*, 64–77. https://doi.org/10.1007/s10979-007-9086-8.

Jurecska, D. E., Peterson, M. A., & Millkey, A. (2012). Comparative study of the MacCAT-CA and CAST-MR for individuals with intellectual disabilities. *American Journal of Forensic Psychology, 30*, 67–89.

Lally, S. J. (2003). What tests are acceptable for use in forensic evaluations? A survey of experts. *Professional Psychology: Research and Practice, 34*, 491–498. https://doi.org/10.1037/0735-7028.34.5.491.

Lipsitt, P., Lelos, D., & McGarry, A. L. (1973). Competency for trial: A screening instrument. *American Journal of Psychiatry, 128*, 105–109. https://doi.org/10.1176/ajp.128.1.105.

Ludecke, D. (2018). Esc: Effect size computation for meta-analysis. R package version 0.4.1 https://CRAN.R-project.org/package=esc.

Marcus, D. K., Poythress, N. G., Edens, J. F., & Lilienfeld, S. O. (2010). Adjudicative competence: Evidence that impairment in "rational understanding" is taxonic. *Psychological Assessment, 22*, 716–722. https://doi.org/10.1037/a0020131.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemica Medica (Zagreb), 22*, 276–282. https://doi.org/10.11613/bm.2012.031.

Menzies, R. J., Webster, C. D., Roesch, R., Jensen, F. A., & Eaves, D. (1984). The fitness interview test: A semi-structured instrument for assessing competency to stand trial, with a proposal for its implementation. *Medicine and Law, 3*, 151–162.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statemen. *Annals of Internal Medicine, 151*, 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135.

Mueller, C., & Wylie, A. M. (2007). Examining the effectiveness of an intervention designed for the restoration of competency to stand trial. *Behavioral Sciences & the Law, 25*, 891–900. https://doi.org/10.1002/bsl.775.

Mumley, D. L., Tillbrook, C. E., & Grisso, T. (2003). Five year research update (1996-2000): Evaluations for competence to stand trial (adjudicative competence). *Behavioral Sciences & the Law, 21*, 329–350. https://doi.org/10.1002/bsl.534.

Neal, T. M. S., & Grisso, T. (2014). Assessment practices and expert judgment methods in forensic psychology and psychiatry. *Criminal Justice and Behavior, 41*, 1406–1421. https://doi.org/10.1177/0093854814548449.

Nicholson, R. A., & Kugler, K. E. (1991). Competent and incompetent criminal defendants: A quantitative review of comparative research. *Psychological Bulletin, 109*, 355–370. https://doi.org/10.1037/0033-2902.109.3.355.

Norton, K. A., & Ryba, N. L. (2010). An investigation of the ECST-R as a measure of competence and feigning. *Journal of Forensic Psychology Practice, 10*, 91–106. https://doi.org/10.1080/15228930903446625.

Otto, R. K., Poythress, N. G., Nicholson, R. A., Edens, J. F., Monahan, J., Bonnie, R. J., ... Eisenberg, M. (1998). Psychometric properties of the MacArthur competence assessment tool–criminal adjudication. *Psychological Assessment, 10*, 435–443. https://doi.org/10.1037/1040-3590.10.4.435.

Pirelli, G., Gottdiener, W. H., & Zapf, P. A. (2011). A meta-analytic review of competency to stand trial research. *Psychology, Public Policy, and Law, 17*, 1–53. https://doi.org/10.1037/a0021713.

Pirelli, G., & Zapf, P. A. (2008). An investigation of psychologists' practices and attitudes toward participation in capital evaluations. *Journal of Forensic Psychology Practice, 8*, 39–66. https://doi.org/10.1080/15228930801947294.

Plonsky, L., & Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. *Modern Language Journal, 100*, 538–553. https://doi.org/10.1111/modl.12335.

Poythress, N. G., Nicholson, R., Otto, R. K., Edens, J. F., Bonnie, R. J., Monahan, J., & Hoge, S. K. (1999). *The MacArthur competence assessment tool-criminal adjudication.* Lutz, FL: PAR.

R Core Team (2013). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Redding, R. (1997). Depression in jailed women defendants and its relationship with their

adjudicative competence. *The Journal of the American Academy of Psychiatry and the Law, 25*, 105–119.

Robey, A. (1965). Criteria for competency to stand trial: A checklist for psychiatrists. *American Journal of Psychiatry, 122*, 616–623. https://doi.org/10.1037/t27252-000.

Roesch, R., Webster, C. D., & Eaves, D. (1984). *The fitness interview test: A method for examining fitness to stand trial toronto.* Canada: Centre of Criminology, University of Toronto.

Rogers, R., Tillbrook, C. E., & Sewell, K. W. (2004). *Evaluation of competency to stand trial - revised (ECST-R).* Odessa, Florida: Psychological Assessment Resources.

Rogers, R., Ustad, K. L., Sewell, K. W., & Reinhardt, V. (1996). Dimensions of incompetency: A factor analytic study of the Georgia court competency test. *Behavioral Sciences & the Law, 14*, 323–330. https://doi.org/10.1002/(SICI)1099-0798(199622)14:3<323::AID-BSL243>3.0.CO;2-Z.

Rogers, T. P., Blackwood, N., Farnham, F., Pickup, G., & Watts, M. (2009). Reformulating fitness to plead: A qualitative study. *Journal of Forensic Psychiatry and Psychology, 20*, 815–834. https://doi.org/10.1080/14789940903174139.

Rutledge, E., Kennedy, M., Neill, H., & Kennedy, H. G. (2008). Functional mental capacity is not independent of the severity of psychosis. *International Journal of Law and Psychiatry, 31*, 9–18. https://doi.org/10.1016/j.ijlp.2007.11.002.

Ryba, N. L., & Zapf, P. A. (2011). The influence of psychiatric symptoms and cognitive abilities on competence-related abilities. *International Journal of Forensic Mental Health, 10*, 29–40. https://doi.org/10.1080/14999013.2010.550982.

Sawilowsky, S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods, 8*, 467–474. https://doi.org/10.22237/jmasm/1257035100.

Scaini, S., Battaglia, M., Beidel, D. C., & Ogliari, A. (2011). A meta-analysis of the cross-cultural psychometric properties of the social phobia and anxiety inventory for children (SPAI-C). *Journal of Anxiety Disorders, 26*, 182–188. https://doi.org/10.1016/j.janxdis.2011.11.002.

Schulze, R. (2004). *Meta-analysis: A comparison of approaches.* Gottingen, Germany: Hogrefe & Huber.

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

Szasz, T. (1988). *Psychiatric justice.* Syracuse, New York: Syracuse University Press.

Ustad, K. L., Rogers, R., Sewell, K. W., & Guarnaccia, C. A. (1996). Restoration of competency to stand trial: Assessment with the Georgia court competency test and the competency screening test. *Law and Human Behavior, 20*, 131–146. https://doi.org/10.1007/BF01499351.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metaphor package. *Journal of Statistical Software, 36*, 1–48. https://doi.org/10.18637/jss.v036.i03.

Viljoen, J. L., Zapf, P. A., & Roesch, R. (2004). Diagnosis, current symptomatology, and the ability to stand trial. *Journal of Forensic Psychology Practice, 3*, 23–37. https://doi.org/10.1300/J158v03n04_02.

Vitacco, M. J., Rogers, R., & Gabel, J. (2009). An investigation of the ECST-R in male pretrial patients: Evaluating the effects of feigning on competency evaluations. *Assessment, 16*, 249–257. https://doi.org/10.1177/1073191108325057.

White, A. J., Meares, S., & Batchelor, J. (2014). The role of cognition in fitness to stand trial: A systematic review. *The Journal of Forensic Psychiatry & Psychology, 25*, 77–99. https://doi.org/10.1080/14789949.2013.868916.

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., ... Bossuyt, P. M. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine, 155*, 529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009.

Wood, M. E., Anderson, J. L., & Glassmire, D. M. (2017). The MacArthur competence assessment tool—Criminal adjudication: Factor structure, interrater reliability, and association with clinician opinion of competence in a forensic inpatient sample. *Psychological Assessment, 29*, 776–785. https://doi.org/10.1037/pas0000368.

Zapf, P. A., & Roesch, R. (2001). A comparison of the MacCAT-CA and the FIT for making determinations of competency to stand trial. *International Journal of Law and Psychiatry, 24*, 81–92. https://doi.org/10.1016/S0160-2527(00)00073-X.

Zapf, P. A., Skeem, J. L., & Golding, S. L. (2005). Factor structure and validity of the MacArthur competence assessment tool—Criminal adjudication. *Psychological Assessment, 17*, 433–445. https://doi.org/10.1037/1040-3590.17.4.433.