Journal of Clinical and Experimental Neuropsychology

(2024) Vol 46(2) 124-140.

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ncen20

Development of a measure for assessing malingered incompetency in criminal proceedings: Denney competency related test (D-CRT)

Robert L. Denney, Sundeep Thinda, Patrick M. Finn, Rachel L. Fazio, Michelle J. Chen & Michael R. Walsh

To cite this article: Robert L. Denney, Sundeep Thinda, Patrick M. Finn, Rachel L. Fazio, Michelle J. Chen & Michael R. Walsh (12 Feb 2024): Development of a measure for assessing malingered incompetency in criminal proceedings: Denney competency related test (D-CRT), Journal of Clinical and Experimental Neuropsychology, DOI: 10.1080/13803395.2024.2314731

To link to this article: https://doi.org/10.1080/13803395.2024.2314731

Published online: 12 Feb 2024.

JOURNAL OF CLINICAL AND EXPERIMENTAL NEUROPSYCHOLOGY
https://doi.org/10.1080/13803395.2024.2314731

Development of a measure for assessing malingered incompetency in criminal proceedings: Denney competency related test (D-CRT)

Robert L. Denney[a], Sundeep Thinda[b], Patrick M. Finn[c], Rachel L. Fazio[d], Michelle J. Chen[e] and Michael R. Walsh[f]

[a]Missouri Memory Center, Citizens Memorial Healthcare, Bolivar, MO, USA; [b]Independent Practice, Surrey, British Columbia, Canada; [c]Department of Mental Health, William S. Middleton Memorial Veterans Hospital, Madison, WI, USA; [d]Independent Practice, Little Rock, AR, USA; [e]Department of Psychological Sciences, University of North Carolina, Charlotte, North Carolina, USA; [f]Departments of Forensic Psychology and Neuropsychology, Burrell Behavioral Health, Springfield, MO, USA

ABSTRACT

Introduction: Experts frequently assess competency in criminal settings where the rate of feigning cognitive deficit is demonstrably elevated. We describe the construction and validation of the Denney Competency Related Test (D-CRT) to assess feigned incompetency of defendants in the criminal adjudicative setting. It was expected the D-CRT would prove effective at identifying feigned incompetence based on its two alternative, forced-choice and performance curve characteristics.

Method: Development and validation of the D-CRT occurred in described phases. Items were developed to measure competency based upon expert review. Item analysis and adjustments were completed with 304 young teenage volunteers to obtain a proper spread of item difficulty in preparation for eventual performance curve analysis (PCA). Test-retest reliability was assessed with 44 adult community volunteers. Validation included an analog simulation design with 101 jail detainees using MacArthur Competency Assessment Test-Criminal Adjudication and Word Memory Test as criterion measures. Effects of racial/ethnic demographic differences were examined in a separate study of 208 undergraduate volunteers. D-CRT specificity was identified with 46 elderly clinic referrals diagnosed with mild cognitive impairment and dementia.

Results: Item development, adjustment, and repeat analysis resulted in item probabilities evenly spread from .28 to 1.0. Test-retest correlation was good (.83). Internal consistency of items was excellent (KR-20 > .91). D-CRT demonstrated convergent validity in regard to measuring competency related information and as well as malingering. The test successfully differentiated between jail inmates asked to perform their best and inmates asked to simulate incompetency (AUC = .945). There were no statistically significant differences found in performance across racial/ethnic backgrounds. D-CRT specificity remained excellent among elderly clinic referrals with significant cognitive compromise at the recommended total score cutoff.

Conclusions: D-CRT is an effective measure of feigned criminal incompetency in the context of potential cognitive deficiency, and PCA is assistive in the determination. Additional validation using knowns groups designs with various mental health-related conditions are needed.

ARTICLE HISTORY
Received 2 October 2023
Accepted 8 December 2023

KEYWORDS
Assessment; malingering; criminal; competency; validity

Malingering has been defined in the DSM-5 as the intentional production of false or grossly exaggerated physical or psychological problems, motivated by external incentives (American Psychiatric Association, 2013). It has been established that malingering can occur across multiple domains: psychological symptoms, cognitive impairment, and somatic complaints (Rogers, 2018), as well as pain (Bianchini et al., 2005). Malingering may take on different forms depending on whether the context is civil or criminal, with civil litigants feigning more cognitive and somatic complaints and criminal litigants feigning more psychopathology (Wygant et al., 2007). There are times, however, when cognitive functioning is pertinent to the issue of competency to proceed, thus setting the stage for potential malingered cognitive dysfunction and ignorance of competency-related knowledge (Denney, 2008a, 2008b, 2012; Denney & Fazio, 2021; Wynkoop & Denney, 1999).

Some commonly used instruments to assist in the assessment of competency include Competency Assessment Instrument (CAI; Laboratory of

CONTACT Robert L. Denney RobertDenneyPsyD@gmail.com Missouri Memory Center, Memorial Healthcare, Bolivar, MO 65613, USA

Note. Portions of the primary research referenced in this manuscript served as portions of doctoral dissertations for the second and third authors as part of their requirements for the Doctor of Psychology (Psy.D.) degree at the School of Professional Psychology at Forest Institute, Springfield, Missouri, USA. The views, opinions, and/or findings contained in this article are those of the authors and should not be construed as an official Department of Veteran's Affairs or U.S. Government position, policy or decision unless so designated by other official documentation.

Community Psychiatry, 1973; McGarry et al., 1973), Revised-CAI (Riley, 1998, March), Georgia Court Competency Test (Wildman et al., 1980), Competence Assessment for Standing Trial for Defendants with Intellectual Disability (Everington & Luckasson, 1992, 2022), MacArthur Competency Assessment Tool – Criminal Adjudication (Poythress et al., 1999), Evaluation of Competency to Stand Trial – Revised (Rogers et al., 2004), and the Fitness Interview Test-Revised (Roesch et al., 2006). None of these instruments have embedded indices to detect feigned ignorance of legal knowledge. The ECST-R has a short scale focused on exaggeration of psychotic and non-psychotic symptoms, but none of those items include content specific to competency. The MacCAT-CA authors recommend concomitant use of a performance validity assessment in order to help identify individuals feigning cognitive dysfunction (Poythress et al., 1999). The above measures also have substantial face validity, providing an easy opportunity for defendants to feign incompetency by simply claiming ignorance of competency-related legal issues (Denney, 2012).

Most techniques designed to identify feigned cognitive impairment (i.e., performance validity tests, PVTs) rely on two strategies to detect feigned deficits. Using a binary response characteristic, they can identify scores that fall statistically below random based on the binomial theorem. That is, the score falls below statistically based expectations for hypothetical individuals with absolutely no cognitive knowledge or ability (defined conceptually as the random range). This range of statistically significantly below random performance has been repeatedly considered the "smoking gun of intent" because it demonstrates the examinee chose the wrong answer more often than what would occur by chance if completely guessing or even blindfolded (Binder & Chafetz, 2018; Pankratz & Erickson, 1990; Slick et al., 1999). The alternative proposal that such a very poor score resulted from the effects of severe impairment is ruled out because it is lower than expectations for individuals with genuine cognitive and/or mental illness who are otherwise providing appropriate task engagement. Research demonstrates that relying on below random performance alone is less sensitive to feigning than criterion-based cutoffs, as most simulators instructed to feign brain impairment, as well as those feigning in known-groups designs, tend to perform better than 50% correct (Ardolf et al., 2007; Binder & Chafetz, 2018; Binder et al., 2014). Consequently, most PVTs incorporate criterion-based cutoff scores developed from known clinical groups in addition to below-random cutoffs.

While there are numerous tests designed to detect feigned cognitive dysfunction, there are few instruments designed to measure feigned incompetency to proceed through criminal adjudication. The Inventory of Legal Knowledge (ILK; Otto et al., 2010) and Test of Malingered Incompetence (TOMI; Colwell et al., 2008) are the only measures available specifically designed for this purpose.

The ILK is a 61-item true/false questionnaire that is administered orally to the criminal defendant using knowledge-based statements about the American legal system. The vast majority of cognitive performance validity techniques rely on two alternative, forced choice (2AFC) testing, whereas the ILK relies on true/false responding. While true/false responding is 2AFC, it is, nonetheless, subtly different than having two seemingly approach-oriented options.

The ILK also uses criterion-based cutoffs developed from psychiatric patients. Guenther and Otto (2010) established criterion-based cutoffs based on simulation designs that included college students ($n = 207$) and patients in short-term residential psychiatric treatment programs ($n = 100$). Those patients were most commonly diagnosed with bipolar disorder (50%), schizophrenia (19%), depression (13%), and some form of anxiety disorder (13%). Both groups were divided into two groups: genuine responding and simulating groups. While the ILK was excellent at identifying simulation for both groups, the specificity rates for psychiatric patients were noticeably lower, with rates of .82–.88 at the same cutoffs established with the college students. These results suggested the ILK may have an unacceptably low specificity in psychiatric populations using standard cutoffs. Interestingly, 3% of the college simulators scored below random, and 20% of psychiatric patients did so. There is little research using the ILK in the context of potential cognitive concerns.

Otto et al. (2011) compared the ILK performance to Rey 15-Item test, Reliable Digit Span, and Test of Memory Malingering among a sample of 120 criminal defendants undergoing competency to proceed examinations. ILK correlated with the measures significantly, having coefficients between .60–.67. No predictive statistics were available. Rogers et al. (2017) evaluated the effectiveness of the ILK using a simulation design among 70 county detainees (50% female) with an estimated intellect in the low average range. The ILK demonstrated sensitivity to simulation of .96 and specificity rate of .89 using standard recommended cutoffs. Requiring below random performance dropped the sensitivity to .43. In another simulation design, but this time with college students, Wahl and colleagues (2018) demonstrated the ILK to have a sensitivity of .67 and

specificity of 1.0 using the standard cutoff. Reviews of the ILK are generally favorable if it is viewed as a screening, at least until test results fall below random, when it can be viewed more definitively (Acklin, 2012; Rubenzer, 2011). Rubenzer suggested the significant need for additional data coming from clinical groups to address potential false-positive rates, particularly cognitively impaired individuals. Indeed, the authors (Otto et al., 2010) note the measure has not been validated for use with defendants having low intellect or significant cognitive deficits, but that below-random performances, nonetheless, demonstrate intent to feign deficits in legal-related knowledge.

Some validation among low IQ individuals has since occurred. Gottfried and Carbonell (2014) considered how the ILK performs among different levels of intellectual efficiency. In addition to college students, they administered the ILK to 65 state psychiatric hospital patients who had been previously adjudicated incompetent to proceed and committed for restoration and who agreed to participate in the study. These individuals typically had multiple psychiatric diagnoses, most commonly psychotic disorders (66.2%) and mood disorders (29.2%). Almost 10% were diagnosed with cognitive disorder or dementia. Intellectual disability (ID) represented .8%, and borderline intellectual functioning comprised .8%. Shipley Institute of Living Scale (Shipley et al., 2009) vocabulary subtest-based IQ estimates varied from 25 to 113 ($M = 69.17$, $SD = 24.3$). None of the detainees were screened for poor effort of invalid performances. Thirteen percent of hospital defendants scored below the standard cutoff on the ILK, but means and standard deviations were not provided. The authors noted ILK total score was moderately correlated to estimated verbal intelligence (.47) when combining college students and forensic patients. Results are limited in their meaningfulness due to the manner in which intelligence was assessed and the lack of screening for malingering.

Watson and Kivisto (2017) administered the ILK to 32 adults with ID who were attending a day program. While all participants were compensated, there was no specific incentive to perform well. Twenty-four of the participants were assigned to a genuine response group, with the remainder comprising a simulation group. The mean Kaufman Brief Intelligence Test, 2nd edition (KBIT-2) IQ of the total sample was 59.79 ($SD = 17.26$). They noted the ILK total score was significantly correlated to KBIT-2 IQ. Mean ILK for the 24 genuine responding participants was 33.96 ($SD = 4.69$). ILK could not differentiate between genuine and simulating groups. Given this sample, the ILK cutoff would have to be lowered to well within the random range for specificity to become acceptable, making the instrument ineffective in this population; however, that conclusion is likely not accurate when evaluating criminal defendants unless they had a history of clear ID diagnosis and involvement in the types of activities typical of such day programming prior to arrest.

At the time of this writing, no validation of the ILK among other cognitively impaired individuals has occurred. Recently, authors recommended possible changes to the ILK to decrease false-positive errors as well as the need to incorporate subscale measures to identify random responding and response style coaching (Danzer & Wasser, 2021). The established cutoffs for the measure appear to have acceptable sensitivity and specificity among psychiatric groups studied thus far, with the possible exception of intellectually disabled individuals, and further research is needed with cognitively impaired individuals.

The Test of Malingered Incompetence (TOMI; K. Colwell et al., 2008) is a 50-item orally administered two alternative, forced-choice questionnaire comprising two 25-item scales designed to detect feigning incompetency to proceed. TOMI-G scale measures simple general knowledge facts about the world, and TOMI-L, which measures simple criminal competency facts. The TOMI-G is administered first, followed by TOMI-L. Originally developed with 72 items on each scale, it was carved down based on performances of 272 university students divided into genuine and simulating groups. The authors kept items with greater than 0.95 probability for correct responding with genuine responders and less than 0.40 correct item probability among simulators. It was then validated alongside TOMM (Tombaugh, 1996) and Rey-15 Item Test (Rey, 1958) among 30 state psychiatric patients, all of whom were committed related to criminal adjudication. TOMI-L correlated more highly to Rey 15 and TOMM than the TOMI-G. Results suggested TOMI-G had transparent face validity, according to the authors. The established cutoff, nonetheless, identified 99% of genuine responders and 86% of simulators on TOMI-G and 100% of genuine responders and 82% of simulators on TOMI-L.

L. H. Colwell and Colwell (2011) administered TOMI alongside the TOMM to 82 inpatient criminal competency restoration cases. Defendants had a variety of psychiatric diagnoses (45.1% psychotic illness, 17.1% mood disorders, and 7.3 with anxiety, somatoform, or adjustment disorders). Nearly 10% were diagnosed with a cognitive disorder as their primary concern. Those performing below standard cutoffs on TOMM Trial 2 and Retention, and believed by the clinical team to be invalid, were classified as feigning. Remaining defendants were classified as valid. Defendants were classified

as feigning psychosis only ($n = 1$), cognitive dysfunction only ($n = 1$), and psychosis and cognitive dysfunction ($n = 9$); 82.9% were considered valid. ROC provided these results: TOMI-G (AUC = .95), TOMI-L (AUC = .95). At recommended cut-offs, TOMI-G demonstrated .86 sensitivity and .92 specificity. TOMI-L demonstrated 1.00 sensitivity and .79 specificity. Four of five defendants diagnosed with ID performed below standard cut off on TOMI-L. Authors suggested the more obvious face validity of TOMI-G may have had the effect of making the TOMI-L appear a more genuine measure of competency for some defendants. Other than the small percentage of cognitive disorders present in the above study, we are not aware of any research using TOMI among neurologic disorders.

Purpose and development of the denney competency related test

This article presents stages of construction for the D-CRT, a two alternative, forced-choice (2AFC) test designed to assess individuals for feigned competency-related knowledge deficits. As noted above, there are only two validity measures specific to CST examinations. There is need for additional validity measures specific to CST because criminal defendants assessed for competency undergo repeat examinations, and court testimony oftentimes reveals the nature of validity instruments to defendants sitting in the courtroom. Under such circumstances, examiners need additional options. The D-CRT was also designed to include performance curve characteristics, unlike most 2AFC tests (e.g., Frederick & Foster, 1991; Frederick et al., 2000).

Performance curve refers to the fact that portions of a measure have differing difficulty levels. For example, there are portions of the Nonverbal Medical Symptom Validity Test (Green, 2008) that are easy and other portions which are relatively more difficult. Reviewing performance across these difficulty levels often allows for differentiating between poor performance caused by exaggeration and poor performance caused by genuine severe cognitive compromise, as one can compare the visual display of that performance to relevant comparison groups. This ability is particularly helpful in contexts where examinees may very well have significant cognitive compromise, like mild cognitive impairment or mild dementia (McGuire et al., 2019). This process can be termed performance curve analysis (PCA) in its simplest terms. The ILK and TOMI do not have performance curves, and it is hypothesized that including differing item difficulties within the D-CRT will enhance its interpretative usefulness through PCA.

The purpose of the D-CRT is not to measure competency to proceed, per se, but rather to develop an instrument that measures feigned incompetence to proceed. The following development stages include item construction, item analysis, reliability analysis, and test validation, including results of predictive statistics and establishing initial optimal cut-off scores. Potential racial and ethnic differences were analyzed. Last, data are included from a sample of cognitively impaired adults to verify effective specificity for established cutoffs. All aspects of this work were successfully reviewed by relevant official human subjects/institutional review boards and adhere to principles outlined in the Declaration of Helsinki.

Method

Phase I: Development

Test items concerning information relevant for a defendant's competency to proceed were constructed from competency to stand trial literature and related case law. Items were created to follow a genuine forced-choice, two alternative, format where the examinee must choose between two alternative options rather than a simple true/false or yes/no format. Of 119 questions created, 114 were adjusted and ultimately chosen after review by an expert panel of four American Academy of Forensic Psychology Fellows with primary expertise in criminal matters. The retained items were judged by reviewers to reflect information relevant to the knowledge domain of competency within the criminal adjudicative process (see Table 1 for a sample of final items). In order to allow for Flesch-Kincaid reading level analysis, each item was reformatted to make a complete sentence.

Table 1. Sample D-CRT items.

In a "plea bargain," the court may
(a) find you guilty of a more serious offense
(a) find you guilty of a lesser offense
A "felony" is
(a) a serious crime
(a) not a serious crime
When someone pleads "guilty" they
(a) are admitting they did the crime
(a) are admitting they saw the crime
"Time served" means
(a) the length of your sentence
(a) the time you have already spend in jail
"Probation" takes place
(a) outside jail
(a) inside jail
The prosecutor asks questions to try to
(a) help you with your case
(a) prove you committed the crime

Note. D-CRT = Denney Competency Related Test.

Flesch-Kincaid analysis revealed the D-CRT to have a grade 7.6 reading level. All statistical analyses were completed using SPSS 20.

Item analysis 1

The test was randomly divided into three 38-item tests to reduce testing time. Eighth-grade students were chosen to complete the tests in order to determine item difficulty because research demonstrates that criminal competency to proceed knowledge begins to falter for ages below 14 years of age (Cowden & McKee, 1995; Grisso et al., 2003; Heilbrun et al., 1996; Sarker, 2011; Savitsky & Karras, 1984). The performances of individuals within this age group would better reveal variability in item difficulty compared to average adults, who would likely find nearly all items quite easy.

Participants

After receiving parental consent, 154 eighth-grade student volunteers were obtained from one public school and two private schools in southern Missouri. The sample included 68 male and 86 female students; the average combined age was 13.79 (*SD* = .47 years). Caucasian students represented 92.2% of the sample; African American, 5.2%; Asian, 1.9%; and Native American, 0.7%.

Procedures

Students were administered the Accuracy Level Test (ALT; Carver, 2000) and one of three portions of the D-CRT. The ALT is a synonym-based reading test that results in a grade equivalency.

Results

The mean reading grade equivalency score of the students from the ALT was 9.9 (*SD* = 2.4). Item probabilities were calculated for all items of the three forms (percent correct), and all 114 items were sorted from lowest probability to highest probability. There were five items that less than 50% of students answered correctly (i.e., < .50 probability), and there were two items all students answered correctly (i.e., 1.00 probability).

In order to improve the spread of item difficulty, it was necessary to increase the difficulty of a percentage of questions. Items were arranged in order of their probability scores and divided into intervals of 10 consecutively ranked items, resulting in 11.4 clusters of items, ranging from lowest to highest percent correct. One question was chosen from each of the lowest 10 clusters to maintain an even increase of difficulty over the entire test. Questions were chosen based on the ease of changing the item to increase difficulty and/or need to clarify confusing items. These 10 modified questions replaced the original items. Flesch-Kincaid readability analysis was again performed on the 114 items resulting in a readability score of 7.6, identical to the first version.

Item analysis 2

The 114 items were again broken down into three 38-item groups to reduce testing time and administered to 150 different eighth-grade students from the same schools after obtaining parental consent.

Participants

This sample included 77 male and 73 female participants. The average age was 13.83 (*SD* = .83). Students were Caucasian (91.3%), African American (5.2%), Asian (1.9%), and Native American (0.7%).

Procedures

Students were administered the ALT (Carver, 2000) and D-CRT. D-CRT item probabilities were again calculated.

Results

Students achieved a mean grade equivalent reading level of 10.2 (*SD* = 2.4). Item probabilities of the 114-item final re-combined D-CRT ranged from .20 to .98 and produced a smooth but rapidly increasing gradient of improved probabilities when all 114 items were aligned according to each respective probability of a correct answer (Figure 1).

Although it has been suggested that items scoring below a probability of .3 should be dropped (Frederick et al., 2000; Hays, 1973), the only item that fell consistently below .3 was item 107 which fell at .22 & .20. The item was reviewed, and it was determined that although the average 8th grader was unlikely to be familiar with a specific term used in the item (Alford Plea), familiarity with the item content was more reasonably expected for those who actually find themselves facing criminal charges (an hypothesis carried over to Phase III). Kuder-Richardson 20 internal consistency results for each of the three partial versions of the test were excellent, ranging from .969 to .976. Significant differences among the mean scores of the three groups were not demonstrated, $F(2, 147) = 2.52$, $p = .084$.

From this portion of the research, it was determined that the D-CRT had a less than optimal, but nonetheless, reasonable reading level (7.6). The measure can also be read to the examinee, allowing for defendants with poor reading skills to complete the test. Items were determined to have acceptable internal consistency and a sufficiently optimal spread of difficulty. A gradient of item difficulty creates

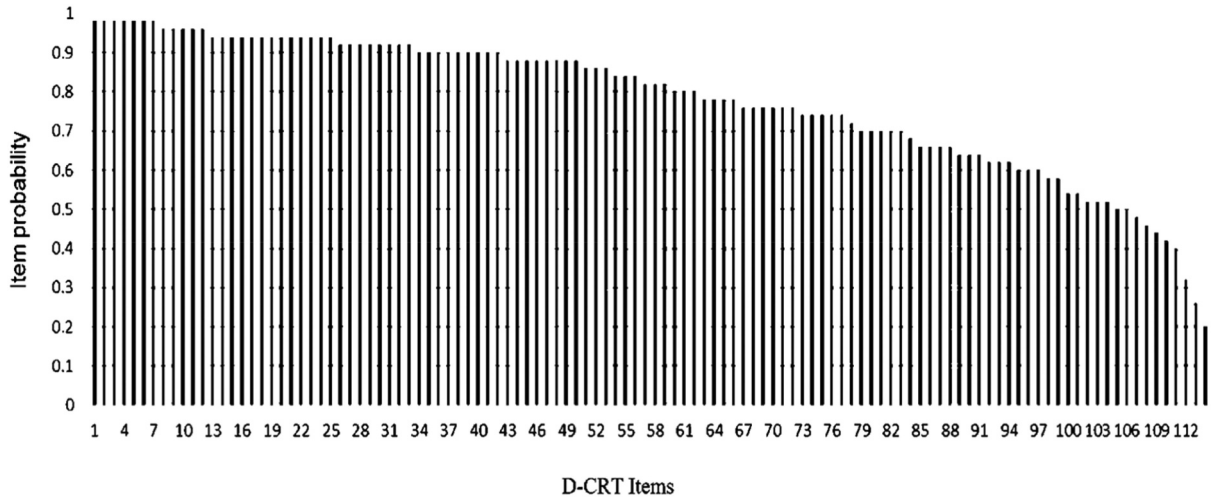

Figure 1. D-CRT item probabilities arranged in order of item difficulty. Note. D-CRT = Denney Competency Related Test. High item probability indicates the item is answered correctly (e.g., .70 = correct 70% of the time). Here we see item difficulty gradually increase across test items.

a performance curve, which has been shown in some measures to enhance the identification of appropriate task engagement (Frederick et al., 2000). The 114 items were rearranged in a random manner, but guaranteeing items from the three development tests were evenly distributed throughout. Random item presentation does not allow the examinee to as easily identify item difficulty.

Phase II: Test reliability

The 114-item D-CRT was administered on two occasions to a convenience sample of 44 adult volunteers in the community. Mean age of the sample was 45.9 (*SD* = 16.5) with 55% female and 45% male. All were Caucasian with none endorsing a specific ethnic background. Mean test-retest interval was 12.23 days (*SD* = 8.63; 1–33 days). Intra-class correlation was calculated using a two-way mixed effects model with absolute agreement, which revealed a high degree of reliability between administrations. The measure ICC was .833 with a 95% confidence interval from .657 to .914, F(43, 43) = 6.92, $p < .001$. Using first assessment data, there were no apparent differences in percent correct between men and women on independent samples t-test, $t(42) = -.083$, $p. = .93$, 95% CI [−2.39, 2.21]. There were age effects among this sample of adults, ages 18 to 74 years, with age improving the scores, $r(42) = .57$, $p < .001$.

Phase III: Test validation

Pilot study

The D-CRT was piloted in a large Midwestern jail for the purpose of demonstrating the item analyses would hold in the 114-item format among the criminal forensic population. Additionally, it was hypothesized that item 107 would perform within acceptable limits in this population.

Participants

The full 114-item version (with items randomized such that they were not in order of item difficulty) was administered to 34 inmates after obtaining informed consent. Inmates were modestly compensated for their participation in the study. The average age of participants was 32.9 years (*SD* = 8.3). Twenty-seven participants were male and seven were female. Regarding race and ethnic self-report, 26 participants were Caucasian, three were Native American, two were African American, two were Caucasian/Hispanic, and one identified as none of the above racial/ethnic groups. The average level of education was 11 years (*SD* = 1.78), and their average reading level was 12.56 (*SD* = 2.26).

Results

It was determined that no changes in item difficulty were needed as all were above the recommended cutoff for exclusion (.3) suggested by Hays (1973) and Frederick et al. (2000). As hypothesized, 70% of inmates answered item 107 correctly. As all of the item probabilities were in an acceptable range, the D-CRT was ready for larger-scale validation.

Main study

The objective of this phase of research was to validate the D-CRT as 1) a test measuring the construct of competency-related knowledge, and 2) a test

measuring the construct of feigning lack of competency to proceed. Again, the goal was not to create a test designed to measure competency to proceed, per se, but rather demonstrate the instrument measures feigned lack of basic legal knowledge. In order to achieve these goals, the D-CRT was administered alongside the McArthur Competency Assessment Tool-Criminal Adjudication (MacCAT-CA; Poythress et al., 1999) and the oral Word Memory Test (WMT, Green & Astner, 1995) in an analog simulation design. The computerized WMT (Green, 2005) was not allowed in the jail. It was expected the D-CRT would correlate positively with the MacCAT-CA and be as effective as the WMT at predicting individuals asked to simulate lack of competency due to cognitive compromise. Simulation designs are useful for initial validation of new instruments and the development of preliminary cutoffs (Heilbronner et al., 2009).

Instruments

The WMT is considered a valid and reliable measure of invalid task engagement based on a verbal learning paradigm (Denney, 2008b; Green, 2005; Green & Astner, 1995; Green et al., 2001; Hartmann, 2002; Wynkoop & Denney, 2005). It has been demonstrated to have 100% sensitivity and specificity in differentiating simulated malingerers from individuals with genuine neurocognitive compromise (Green, 2005).

The MacCAT-CA is a valid and reliable measure of competency to stand trial (Denney, 2012; Poythress et al., 1999; Rogers, 2007). It consists of 22 items that load on three constructs: Understanding, Reasoning, and Appreciation. Questions for Understanding and Reasoning sections revolve around a vignette that is read at the beginning of the test regarding two men who get into a bar fight. The Understanding section measures such things as roles of lawyers, judge, and jury, as well as understanding charges and basic courtroom proceedings. The Reasoning section measures strategy, consequences, conceptualization of mitigating circumstances and information, as well as ability to perform a cost-benefit analysis. The items in Appreciation cover examinees' own case using the story at the beginning of the testing as a comparison. The Appreciation portion was not administered because it was against jail policy for mental health staff and volunteers to ask about inmate's specific legal cases. Each section is scored individually with these score meanings: 11–16 minimal or no impairment; 9–10 mild impairment; and < 9 clinically significant impairment. There is no total score.

The D-CRT as the experimental measure was used in its full 114-item pencil and paper format (it can also be administered orally or on a computer platform). Administration time was approximately 30 minutes. Because the D-CRT was developed using three alternate shorter forms, there were several items that duplicated the same information nearly exactly or by a slight rewording. As a result, a consistency index was developed using 16 item pairs spread evenly throughout the test. It was expected the consistency index would function either as a verification of test comprehension or as an additional index of potential feigning.

Participants

New participants were recruited from the same jail; $N = 103$ volunteers agreed to participate in this portion of the study. Two participants were excluded from data analysis because one was observed answering in what appeared to be a random fashion and completed the task under 10 minutes; the other one chose not to finish the testing. Participants were again modestly compensated for participating.

One hundred one participants completed the tasks after random assignment to groups: 51 participants were designated as Normal Control (NC) group and the remaining 50 were asked to participate as Simulated Malingerers (SM). The overall demographics for the sample revealed a mean age of 30.7 years ($SD = 9.0$). Participants were made up of 67 males and 34 females; 72 participants were Caucasian, 21 African-American, four Native American, and four Other. Mean educational level was 11.3 years ($SD = 1.7$). Statistically significant differences were not demonstrated between groups on age, $t(99) = .486$, $p = .63$, 95% CI [−2.70, 4.45] or education $t(99) = -.79$, $p = .434$, 95% CI [−.93, .40]. Statistically significant differences were also not demonstrated between groups on variables of self-identified race, X^2 (3, $N = 101$) = 2.24, $p = .53$ or gender, X^2 (1, $N = 101$) = 0.834, $p = .40$.

Thirty-six participants (35.6%) reported histories of head injury, including one who reported a resulting six-month coma. The other head injuries either had unknown durations of loss of consciousness (LOC) or were reported as minutes to an hour. One participant reported 5 head injuries with LOC of less than an hour for each. Finding about 36% of jail detainees reporting histories of head injury is consistent with estimates of self-reported head injuries within the criminal legal system based on a recent literature review and meta-analysis (Hunter et al., 2023). Seven participants from the NC (13.7%) reported neurological/psychiatric diagnoses with the following conditions: Bipolar disorder, seizure disorder, short-term memory loss, Schizophrenia, "blood clot on the brain," traumatic brain injury, and one case of Dissociative Identity

Disorder. Three of the SM (6%) group reported neurological/psychiatric diagnoses with the following conditions: spinal cord injury, hearing voices, gunshot wound to the head, and spina bifida. Self-reported medical histories were not verified.

Procedures

All participants were administered the tasks in the following order: Word list presentation trials and Immediate Recognition of the oral WMT, Mac-CAT-CA (Understanding and Reasoning portions), WMT Delayed Recognition, and finally, D-CRT. These instructions were provided to the NC group:

> Today you will be given a memory test, a competency test, and a test being studied by the researchers. It is important that you do your very best on all of the items. Listen carefully to the instructions and answer each question as best you can.

The SM group was provided these instructions:

> You have been assigned to the experimental group. Your job is to act as if you are not competent to go to trial. Try to answer the questions like someone who has had a major brain injury or someone who has mental retardation. Imagine your ability to convince me that your being not competent will result in the dropping of charges against you. You will be given a memory test, a competency test, and a test being studied by the researchers. Do your very best to answer in a way that will make me think that you shouldn't go to trial.

Performance on the WMT in the NC group was used as a verification of proper task engagement. Scores on the WMT were not used as an indicator of effort in the SM group, as low scores were considered reflective of response strategy.

Results

Although the Mac-CAT-CA has well-defined rules by which the instrument should be scored, there is room for error, nonetheless, given the wide variety of potential responses an examinee can provide. Ratings of MacCAT-CA items were performed by two researchers (PMF & RLF) blinded to group membership and each other's ratings. Correlations between the two raters for Understanding and Reasoning sections were high, N =101, r_s =.91, p < 001, 95% CIs [.84, .96] and r_s = .84, p < .001, [.71, .92], respectively. Inconsistencies between ratings were blindly resolved by RLD, resulting in the final scores used in analyses.

Obtained scores for both groups across measures are presented in Table 2. Due to finding unequal variances for all scores between groups, non-parametric statistics were used. Scores obtained by SM were significantly poorer than results obtained by NC on all measures. Mean WMT scores for NC fell within the valid range based on standard WMT cutoffs, while mean WMT scores for SM fell within the invalid range. WMT findings supported the belief that participants largely followed their respective directions. The finding that MacCAT-CA scores were significantly lower for SM than NC verified that individuals simulated lack of criminal competence. D-CRT total score was significantly lower for SM group compared to NC. It also appeared that simulators were less consistent in their responding throughout the D-CRT, suggesting consistency was at least a part of their simulation strategies. The effect sizes for D-CRT percent correct and internal consistency variables were large. Internal consistency of the D-CRT was good to excellent among the 51 genuine controls, KR-20 (N = 102) = .917, 95% CI [.881, .956].

Significant correlations were also found between the D-CRT and the other measures (Table 3). Item

Table 2. Statistical results of measures by group for the jail sample.

	Controls		Simulators			Hodges-Lehman Estimate	
Measures	*MSD*	*SD*	*M*	*SD*	*U*	95% CIs	$r = \frac{\lvert z \rvert}{\sqrt{n}}$
WMT IR	94.51	4.98	58.90	26.72	266.00	35.00	.69
					z = 6.88	25.00, 45.00	
WMT DR	96.67	5.99	59.60	27.84	255.00	35.00	.70
					z = −7.0	27.50, 45.00	
WMT CNS	93.92	6.07	71.85	18.10	344.50	22.50	.63
					z = −6.36	15.00, 30.00	
MacCAT-CA Understanding	12.14	2.26	4.00	3.40	69.50	8.00	.82
					z = −8.21	7.00, 10.00	
MacCAT-CA Reasoning	13.78	1.87	7.54	4.33	199.50	6.00	.73
					z = −7.33	4.00, 7.00	
D-CRT Total Correct %	88.96	8.81	58.16	15.14	141.50	32,46	.77
					z = −7.70	28.07, 36.84	
D-CRT Consistency Index %	93.98	8.74	61.04	16.84	129.00	32.00	.79
					z = −7.89	31.00, 38.00	

Note. N = 101. WMT = Word Memory Test. IR = Immediate Recognition percent correct. DR = Delayed Recognition percent correct. CNS = Consistency between WMT IR & DR percent correct. MacCAT-CA = MacArthur Competency Assessment Tool-Criminal Adjudication. D-CRT = Denney Competency Related Test. All U p < .001.

Table 3. Descriptive statistics and correlations for jail study variables.

Measure	*M*	*SD*	1	2	3	4	5	6	7
1. WMT IR	76.88	26.13	–						
2. WMT DR	78.32	27.29	.91	–					
3. WMT CNS	82.10	17.38	.72	.67	–				
4. MacCAT-CA Understanding	8.11	4.99	.71	.70	.62	–			
5. MacCAT-CA Reasoning	10.69	4.56	.63	.65	.58	.81	–		
6. D-CRT Total	73.71	19.76	.63	.66	.62	.73	.64	–	
7. D-CRT CNS	77.67	21.24	.58	.59	.60	.67	.66	.79	–

Note. N = 101. WMT = Word Memory Test. IR = Immediate Recognition. DR = Delayed Recognition. WMT CNS = Consistency between WMT IR & DR. MacCAT-CA = MacArthur Competency Assessment Tool-Criminal Adjudication. D-CRT = Denney Competency Related Test. D-CRT Total = Total percentage correct. D-CRT CNS = D-CRT internal consistency. All correlations $p < .001$.

probabilities were analyzed and ranged from .28 to 1.0 for the NC group and .28–.82 for the SM group. Item probabilities only rose to .82 among the simulation group, suggesting simulators tended to perform poorly on a greater number of easier items.

Predictive statistics were performed to identify how well D-CRT identified individuals simulating lack of competence. According to Hosmer and Lemeshow (2000), Receiver Operating Characteristics Area Under the Curve (AUC) of .700–.799 is acceptable, .800–.899 is excellent, and .900–1.0 is outstanding. AUCs of WMT subscales demonstrated excellent to outstanding utility at predicting simulation (IR = .896, $p < .001$, 95% CI [.826, .966]; DR = .90, $p < .001$, 95% CIs [.835, .965]; CNS = .865, $p < .001$, 95% CI [.792, .938]). D-CRT percentage correct was also outstanding at predicting simulated incompetence (AUC = .945, $p < .001$, 95% CI [.898, .991], Figure 2).

Sensitivity and specificity rates for the D-CRT are presented in Table 4 at various total percent correct cutoff scores. Choosing a cut score that maintains a high specificity of at least .90 (i.e., false-positive rate at or below .10) is important when developing measures of performance and symptom validity (Boone, 2007). Using this standard results in a cutoff score of ≤ 83% and establishes a sensitivity to

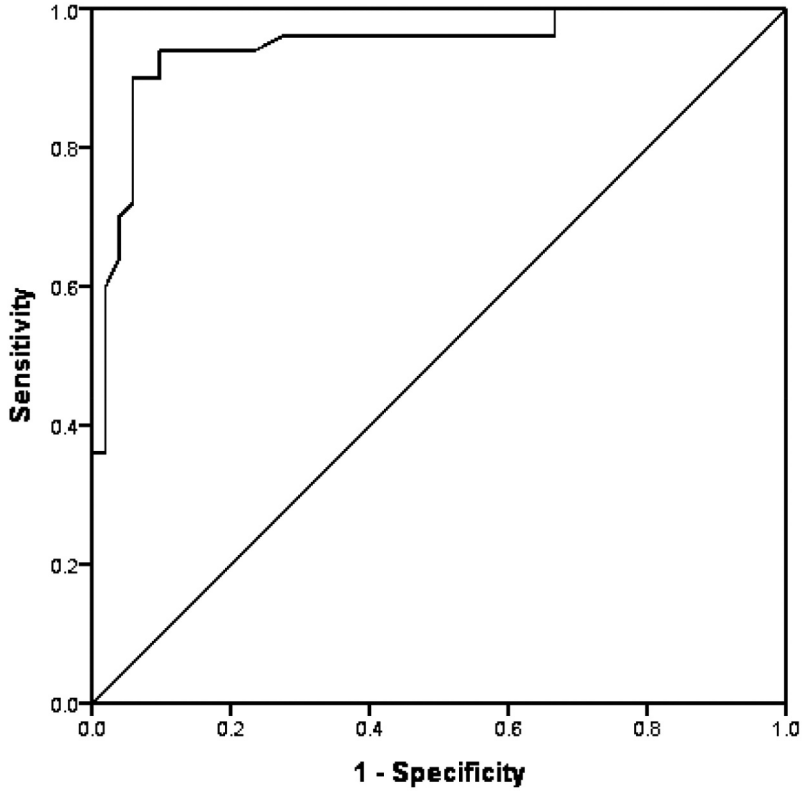

Figure 2. AUC of D-CRT predicting simulators. Note. N = 101. AUC = Area Under Curve. D-CRT = Denney Competency Related Test. Corner to corner diagonal line represents no predictive utility. Short diagonal lines represent tied scores on the D-CRT.

Table 4. Sensitivity and specificity rates for the D-CRT at various cutoffs for the detection of simulated incompetency among adult jail detainees.

Positive if Less Than or Equal To (%):	Sensitivity	Specificity
51	.360	1.000
52	.360	0.980
53	.400	0.980
54	.420	0.980
56	.460	0.980
57	.500	0.980
58	.520	0.980
59	.560	0.980
60	.600	0.980
61	.640	0.961
62	.680	0.961
64	.700	0.961
66	.720	0.941
67	.740	0.941
68	.760	0.941
69	.780	0.941
71	.800	0.941
72	.820	0.941
75	.860	0.941
77	.880	0.941
79	.900	0.941
81	.900	0.902
82	.920	0.902
83	.940	0.902
84	.940	0.882
85	.940	0.843
86	.940	0.765
87	.960	0.725
88	.960	0.667

Note. N = 101. D-CRT = Denney Competency Related Test.

simulated incompetence of 94%. Using a cutoff score of ≤ 79 resulted in sensitivity of .90 while increasing the specificity to .94 (6% false-positive rate). Percent correct cutoffs for statistically rare performances using a one-tail standard (Siegel & Castellan, 1988) are 43% at p = .08, 41% at p = .04 and 39% at p < .01. Twelve percent of jail simulators performed below random at $p \leq .05$.

Phase III: Test reliability

The 114-item D-CRT was administered on two occasions to a convenience sample of 44 adult volunteers in the community. Mean age of the sample was 45.9 (SD = 16.5), with 55% female and 45% male. All were Caucasian, with none endorsing a specific ethnic background. Mean test-retest interval was 12.23 days (SD = 8.63; 1–33 days). Intra-class correlation using a two-way mixed effects model with absolute agreement revealed a result that was moderate to excellent at .83 (95% CI [.66, .91] for percent correct. Using first assessment data, there were no apparent differences between men and women on independent samples t-test, $t(42) = -.083$, p = .934, 95% CI [−2.40, 2.21]. There were age effects among this sample of adults (ages: 18–74 years), with age improving the scores, N = 44, r = .57, p < .001, 95% CI [.321, .738]. The mean consistency index for this first assessment data was 97.60 (SD = 3.61).

Phase IV: Racial/ethnic demographic effects

Participants recruited from a large southeastern public university completed the D-CRT administered via online survey. Of the total 225 respondents, 11 were removed due to missing data and six were removed because their total scores were not above the random range. The final sample consisted of 208 students. The sample included 145 females (69.7%), 53 males (25.5%), and 10 identifying as non-binary/prefer not to say (4.8%). Mean age of this sample was 19.40 (SD = 2.65), and most respondents were first- or second-year students (n = 107, 51.4%; n = 55, 26.4%, respectively). Only one individual endorsed Native Hawaiian or Pacific Islander, so the score was included under Multiracial or Other. The racial/ethnic breakdown was as follows: White (n=94, 45.2%), Black (n = 43, 20.7%), Latine (n = 18, 8.7%), Asian (n = 26, 12.5%), and Multiracial or Other (n = 27, 13.0%). Nineteen participants reported having personal experience with the criminal justice system (9.1%), and most endorsed enjoying crime-related entertainment (n = 163, 78.4%). The mean D-CRT percent correct score for all participants was 84.87% (SD = 6.87%), and mean consistency was 90.5% (SD = 9.30). Multivariate analysis of variance (MANOVA) revealed no statistically significant differences between groups for total correct ($F(5, 202) = 1.67$, p = .143, ηp^2 = .04) or consistency ($F(5, 202) = 1.11$, p = .356, ηp^2 = .03). Internal consistency of the D-CRT was very high (KR-20 = .967). These findings demonstrated no significant difference in D-CRT performances between minority groups.

Phase V: Specificity among cognitively impaired

The D-CRT was administered to patients from a local healthcare system who volunteered to participate in research. All patients with Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) scores between 28 and 12 during the research period were asked to be involved in research and forty-six patients consented to participate. Informed consent was provided for each patient as all were their own guardians. Each patient was undergoing assessment and treatment as part of routine neurological or psychiatric care either at an outpatient memory center or as part of their hospitalization within an inpatient geropsychiatry unit. All patients completed the MoCA. MoCA is a screening test that has been found reasonably sensitive and

specific for identifying neurocognitive compromise among elderly individuals (Smith et al., 2007). Scores < 26 are generally viewed to reflect cognitive impairment, with ≤ 18 generally viewed as consistent with dementia (Trzepacz et al., 2015).

Patient volunteers were offered the option to complete the D-CRT in one of three ways: Orally to the examinee, within a computer software-guided interface requiring reading and answer selection, or pen and paper booklet. One patient used the computer software (2.2%); 12 requested oral administration (26.1%); and 33 read the written items and marked their answers on test booklets (71.7%). Two patients within the inpatient geropsychiatry unit with MoCA scores of 17 consented but were not able to answer even oral D-CRT questions (they became confused with the items or were unable to make selections). One patient with a dementia diagnosis (MoCA = 18) within the outpatient clinic started the D-CRT but chose not to complete the test.

Diagnoses were extracted from the electronic health record for the final sample of 46 patients (18 women/28 men). Patient ages spanned 50 to 88 years, with a mean of 74.28 years (*SD* = 8.12). Their educations ranged from 8 years to 20 years (*M* = 13.93, *SD* = 3.47). MoCA scores for this sample ranged from 12 to 28 (*M* = 20.91, *SD* = 3.49). Although two MoCA scores were above 25, all cases were either diagnosed as Mild Cognitive Impairment or dementia (Table 5). Mean D-CRT correct for all-cause MCI (*N* = 23) was 90.50% (*SD* = 3.08). Mean MoCA score for those with all-cause dementia diagnoses (*N* = 23) was 18.61 (*SD* = 2.81). Mean D-CRT correct for those with all cause dementia was 87.45% (*SD* = 5.10). The dementia mean score was within one standard deviation of the test-retest volunteers and normal controls in the jail sample. Regarding D-CRT correct, no MCI patient scored below 84%, and no dementia patient scored below 79%. No significant difference was identified between written and oral administration for total score percent correct, $t(43) = 1.51$, $p = .139$, 95% CI [−.75, 5.22] or consistency, $t(43) = .307$, $p = .760$, 95% CI [−4.45, 6.05].

Kruskal-Wallace test analysis between memory-disordered patients and the jail sample (*N* = 139) revealed a significant difference on D-CRT percent correct only when compared to simulators $\chi^2(2) = 77.92$, $p < .001$ (see Table 6 for pairwise comparisons). There were no significant differences demonstrated between memory-disordered patients and the genuine jail sample. Figure 3 reveals a generalized separation between jail SM and the other two groups, with the exception of three outlier cases among the jail CN group. Last, there was no apparent significant difference in D-CRT consistency index between memory-disordered patients and the genuine jail sample, $U(89) = 786.5$, $p = .110$, $r = 0.17$, Hodges-Lehman estimate = 0, 95% CI [0.0, 6.20]. These data suggest that information required for successful performance and consistency on the D-CRT is not lost in most dementing cognitive conditions until the condition approaches the severe (i.e., untestable) range.

Table 5. Patient test scores by diagnostic category.

		MoCA		D-CRT Correct %		D-CRT Consistency %	
Diagnoses	*N*	*M*	*SD*	*M*	*SD*	*M*	*SD*
MCI (undefined)	15	22.87	2.80	91.40	2.78	95.02	6.75
MCI amnestic	2	25		89.91	0.62	93.75	8.84
MCI & Bipolar I disorder	1	25	–	90.35	–	100	–
MCI NPH	1	25		86.84		93.80	
MCI vascular	4	22.75	1.26	88.38	0.32	92.23	3.15
All-Cause MCI	23	23.22	2.43	90.50	3.08	94.59	6.04
Dementia (undefined)	1	19	–	93.37	–	100	–
Dementia AD	15	18.87	2.80	87.08	4.20	89.61	8.73
Dementia vascular	4	22.75	1.25	88.38	4.32	92.23	3.15
Dementia mixed	2	19	0.0	83.77	5.58	96.90	4.38
Lewy Body Dementia	2	17	0.0	88.16	8.06	90.65	13.22
All-Cause Dementia	23	18.61	2.81	87.45	5.10	90.79	8.60

Note. *N* = 46. MoCA = Montreal Cognitive Assessment; D-CRT = Denney Competency Related Test; MCI = Mild Cognitive Impairment; NPH = Normal Pressure Hydrocephalus; AD = Probable Alzheimer disease.

Table 6. Dunn's *post hoc* pair-wise comparisons of D-CRT total score between memory disordered patients and jail sample.

Sample 1-Sample 2	Dunn's value	Standard Error	Standard Test Statistic	*p*	η^2
Jail SM-Memory Clinic	−58.97	8.66	−6.81	<.001	.336
Jail SM-Jail NC	65.25	8.01	8.15	<.001	.481
Memory Clinic-Jail NC	6.28	8.62	0.73	=.47	–

Note. *N* = 139. SM = Simulators, NC = Jail genuine controls, D-CRT = Denney Competency Related Test.

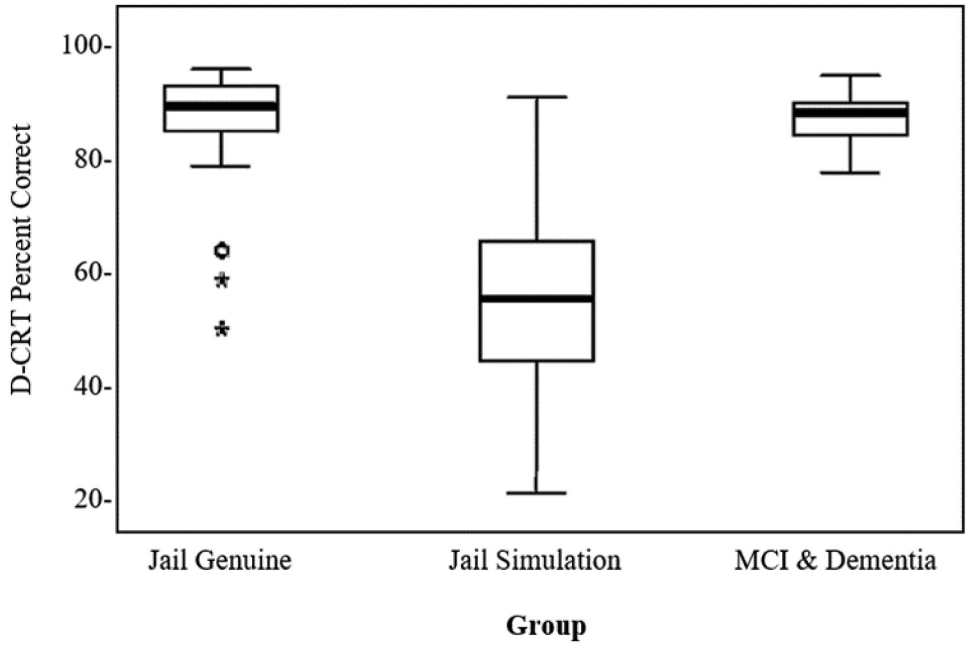

Figure 3. Jail and memory clinic D-DRT percent correct. Note. D-CRT Denney Competency Related Test. NC = Genuine Controls. SM = Simulators. MCI = Mild Cognitive Impairment. Symbols (o) and (*) represent outlier cases.

Table 7. Positive predictive power and negative predictive power given demonstrated D-CRT sensitivity to simulation and specificity against genuine substantive cognitive compromise using recommended cutoff score.

Base Rates	.30	.40	.50
Positive Predictive Power	.774	.842	.889
Negative Predictive Power	1.00	1.00	1.00

Note. D-CRT = Denney Competency Related Test.

Predictive Utility Among Differing Base Rates

We combined the jail simulation group with the memory clinic sample to create a hybrid known groups, simulation study. This method improves the generalizability by incorporating individuals with known and substantial neurocognitive impairment when identifying positive predictive power (PPP) and negative predictive power (NPP) based on presumed base rates of malingered neurocognitive dysfunction. The method remains weak in regard to its tendency to overestimate PPP because simulators theoretically lack the motivation to remain unidentified like genuine defendants who are malingering. However, we suggest the method results in robust NPP. Using the recommended cutoff (≤ 83, Table 3) with these groups resulted in the expected sensitivity of .90 and specificity of 1.0. PPP and NPP at three example base rates are provided in Table 7.

Discussion

We reviewed the development and initial validation of the D-CRT. The goal was to develop a true forced-choice measure based on knowledge content from within the construct of criminal competency to proceed that included aspects of a performance curve. Items were developed from competency to stand trial literature with expert review and item probabilities were examined, adjusted, and verified based on understanding among community young people. Results demonstrated items have a probability spread conducive to the potential development of performance curve analysis. Reading level of items consistently fell within the 7th grade. While this reading level may be above that of some of criminal defendants, the D-CRT can be administered verbally without trouble, as evidenced by successful oral administrations in the memory clinic. Internal consistencies were excellent and all above .90. Test-retest reliability was moderate to good with an overall ICC of .83. There were no racial/ethnic demographic effects demonstrated on D-CRT percent correct or consistency scores among undergraduate volunteers.

Among a jail population, the D-CRT demonstrated convergent construct validity regarding criminal competency based on statistically significant positive correlations with the MacCAT-CA. D-CRT correlated slightly higher for the Understanding portion of the

MacCAT-CA compared to the Reasoning portion, which makes sense as D-CRT items were derived from competency-related knowledge based on published literature and case law. Because competency to proceed requires more than just knowledge, the D-CRT is not a competency assessment tool, per se; however, it is by design a basic measure of factual competency-related knowledge.

In a simulated study among jail detainees, D-CRT's statistically significant positive correlations with the WMT validity subtests demonstrated convergent validity regarding simulated malingering. The D-CRT demonstrated outstanding predictive utility in identifying simulated malingering with an AUC of .945. Last, the D-CRT appeared to display excellent specificity at the recommended cutoff score when administered to adults with clinically significant cognitive compromise (i.e., apparent low false positive rate). Given these results, it appears that cognitive deficits alone, even to the point of moderate dementia, would not explain below cutoff score performances on the D-CRT.

We have limited data regarding the effects of active psychosis on the D-CRT. Although nearly 14% of the genuine sample within the jail validation sample included individuals with self-reported mental illnesses, they were presumably not acutely ill as they were housed in the general population. Additionally, significant affective disorders were present in a small number of memory clinic patients. It is reasonable that paranoid delusions focused on criminal charges or courtroom participants might alter D-CRT results. It is unclear how major mental illness would eliminate basic knowledge about the criminal justice system if examinees are not significantly distracted from internal stimuli or unwilling to participate in the examination. It is important to note that such difficulties should also present outside the examination if they are severe enough to alter test results. Ultimately, effects of major mental illness on the D-CRT is an empirical question. The purpose for developing the D-CRT, was to increase validity assessment options specific to criminal competencies when significant cognitive compromise is the referral concern.

There were age effects demonstrated during the first administration of test-retest study; however, those effects were not demonstrated among the genuine sample of jail participants, even though their ages ranged from 17 to 56. The age effects demonstrated in the reliability study may have been due to three subjects with ages approximating 20 who appeared to perform more poorly than the majority of other participants. However, the age effect was also present during the retest portion of the study, although less pronounced. The age effect finding may have been an artifact within that volunteer sample. It may also represent the increased accumulated knowledge regarding the legal system that occurs as one ages (presumably from television programs, movies, and crime novels). This effect may not have been apparent in the jail sample because of the experience jail detainees already had with the criminal justice system. Consequently, that effect may not arise in forensic practice context unless one is dealing with a young defendant who has had no prior interaction with the justice system. Additional data is needed to verify potential age effects.

Initial D-CRT findings compared favorably to the ILK and TOMI. The D-CRT is longer than both of those measures, which is known to generally add measurement stability, and in the case of forced-choice, two alternative tests, makes the measure more sensitive to below-random performance (Siegel & Castellan, 1988). Unlike the ILK true/false format, the D-CRT uses true two alternative choice items. ILK sensitivity in the Rogers and colleagues (2017) study of jail simulators was 0.96, while D-CRT sensitivity in the similar population was 0.90. The ILK specificity was 0.89 (11% false-positive rate) in jail detainees, while D-CRT specificity at the recommended cutoff was higher at 0.941 with jail detainees (6% false-positive rate). The ILK has no data with patients who have mild to moderate cognitive impairment, whereas D-CRT appears to have very low false-positive rates among cognitively impaired older patients.

While the TOMI has 50 items, only 25 actually pertain to legal knowledge. At the time of this writing, TOMI is the only one of these measures to have data from actual known groups design among criminal defendants hospitalized for restoration purposes (10% of whom had cognitive concerns as their primary diagnoses). TOMI results were only reported individually for TOMI-G and TOMI-L. Based on TOMM performance (and clinical judgment) as the criterion predictor of malingering, TOMI-G and -L had good sensitivities at 0.86 and 1.00, respectively, but only TOMI-G had acceptable specificity (0.92). TOMI-L specificity was 0.79 (21% false-positive rate). D-CRT results compare favorably to TOMI predictive statistics, particularly when substantive cognitive concerns are present. Finally, the ILK and TOMI do not have variable item difficulty amenable to a performance curve analysis.

Performance curve within the D-CRT warrants further discussion. Item difficulties are spread evenly across the instrument, but not in a progressively

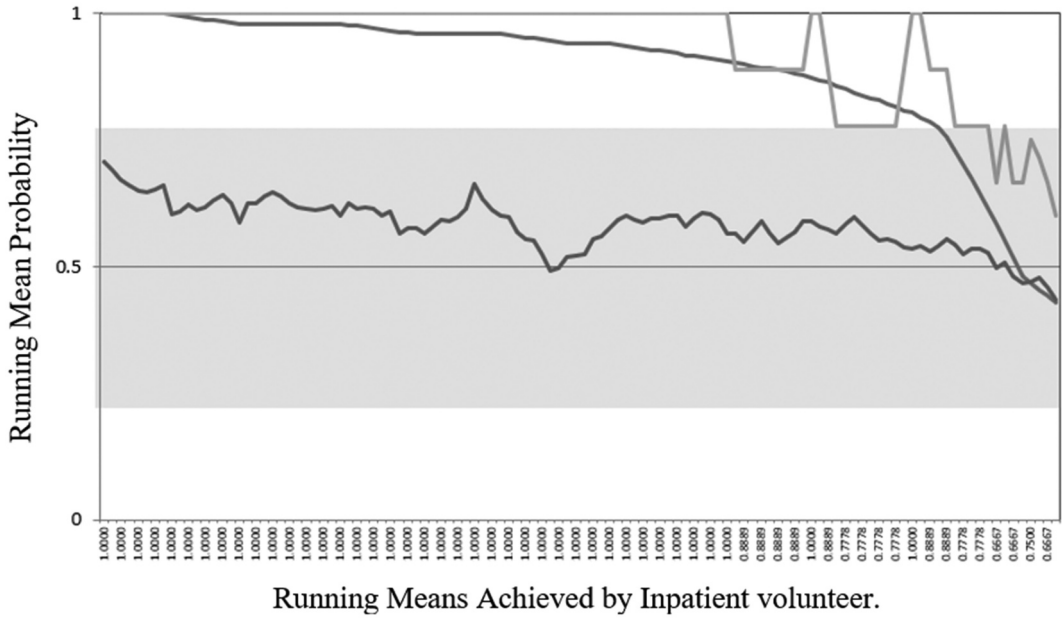

Figure 4. Example D-CRT performance curve. Note. D-CRT = Denney Competency Related Test. Running means have probability between .00 and 1.0 (with 1.0 indicating perfect performance on that set of 10 items). The gray band across the test reflects the boundaries of random responding based on the binomial theorem. The jagged line in the upper right is the performance of a 65-year-old volunteer from an inpatient geropsychiatry ward, who had diagnoses of schizophrenia, bipolar disorder, and dementia. The smooth curve reflects the average performance of jail volunteers asked to perform their best, and the mountain range line reflects average performance of jail volunteers simulating incompetency.

increasing manner. Consequently, examinees, theoretically, have added difficulty in determining which items are more difficult. We have created software that realigns the item results in order of item difficulty from easy to difficult. The software then creates a series of running means across the entire test based on uniform sets of items. In other words, items 1–10 are used to calculate the first mean, and 2–11 create the second mean, and so on across the entire test. This general method was developed by Frederick (Frederick & Foster, 1991; Frederick et al., 2000). These running means are then plotted. This graph creates a picture of the examinee's performance across the test. We found that simulators did not produce a normal performance curve like the genuine jail volunteers did. This graph is helpful in determining if an examinee performance appears as it should. Figure 4 displays the performance curve of a 65-year-old woman with 8 years of education who was a volunteer from within the inpatient geropsychiatry ward. Her diagnoses from the electronic medical record included paranoid schizophrenia, bipolar disorder, and dementia. She completed the test orally. She performed well above the cutoff for total correct and produced a reasonably high consistency index. She produced the jagged line in the upper-right corner of Figure 4 that reveals perfect (100%) performance for the easy two-thirds of the test, with her score falling off during the last quarter. The pattern more closely approximates the smooth average curve of the genuine jail volunteers rather than the simulators. Simulators more commonly produce profiles that simulate a mountain range. While informative, additional research is needed to develop mathematical rules to best benefit from the performance curve.

Construct validation of the D-CRT was done within an analog simulation design. While these designs have merit during the developmental phases of an instrument, they are weak in terms of external validity (Heilbronner et al., 2009). The measure needs cross-validation using known groups designs in order to verify its predictive utility; furthermore, such designs can better establish optimal cutoffs among specific clinical groups (Sweet et al., 2021).

Jail detainees used in this study were from rural southwest Missouri. The average education in the jail sample was just under 12th grade. Given the population served by this jail, it is fair to presume also that a large percentage of these individuals were from very rural and poor areas around the Missouri Ozarks. In many respects, the genuine effort group was a reasonable representation of typical criminal defendants. The small sample size, particularly for the cognitive impairment sample, limits generalizability to some degree.

Further validation studies, particularly with known group designs among those with major mental illness and additional cognitively impaired populations need to occur to adjust cutoff scores and cross-validate the utility of the D-CRT.

Additional research is needed to develop the PCA algorithm and evaluate its effectiveness over and above the percent correct and consistency scores and simply a visual display. Results of this initial research indicate the D-CRT is a reliable and valid measure of feigning lack of criminal competency.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Robert L. Denney http://orcid.org/0000-0002-1171-7248
Rachel L. Fazio http://orcid.org/0000-0001-8996-5324

Author contributions

Robert L. Denney formulated and directed the overall study conceptualization, development, and design with assistance, during the latter phases, of Rachel L. Fazio. Material preparation, data collection, and analyses were performed equally by Sundeep Thinda, Patrick M. Finn, Rachel L. Fazio, Michelle J. Chen, and Robert L. Denney. Software programming was performed by Michael R. Walsh. The first and final drafts of the manuscript were written by Robert L. Denney with authors commenting on various versions of the manuscript.

References

Acklin, M. W. (2012). The forensic clinician's toolbox I: A review of competency to stand trial (CST) instruments. *Journal of Personality Assessment*, *94*(2), 220–222. https://doi.org/10.1080/00223891.2011.627970

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Author.

Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *The Clinical Neuropsychologist*, *21*(6), 899–916. https://doi.org/10.1080/13825580600966391

Attix, D. K., Donders, J., Johnson-Greene, D., Grote, C. L., Harris, J. G., & Bauer, R. M. (2007). Disclosure of neuropsychological test data: Official position of division 40 (clinical neuropsychology) of the american psychological association, association of postdoctoral programs in clinical neuropsychology, and american academy of clinical neuropsychology. (clinical neuropsychology) of the american psychological association, association of postdoctoral programs in clinical neuropsychology, and american academy of clinical neuropsychology. *The Clinical Neuropsychologist*, *21*(2), 232–238. https://doi.org/10.1080/13854040601042928

Bianchini, K. J., Greve, K.W., & Glynn, G. G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *The Spine Journal*, *5*(4), 404–417. https://doi.org/10.1016/j.spinee.2004.11.016

Binder, L. M., & Chafetz, M. D. (2018). Determination of the smoking gun of intent: Significance testing of forced choice results in social security claimants. *The Clinical Neuropsychologist*, *32*(1), 132–144. https://doi.org/10.1080/13854046.2017.1337931

Binder, L. M., Larrabee, G. J., & Millis, S. R. (2014). Intent to fail: Significance testing of forced choice test results. *The Clinical Neuropsychologist*, *28*(8), 1366–1375. https://doi.org/10.1080/13854046.2014.978383

Boone, K. B. (2007). A reconsideration of the slick et al. (1999). Criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 29–49). Guilford Press.

Carver, R. P. (2000). *Technical manual for the accuracy level test*. Revrac Publishing.

Colwell, L. H., & Colwell, K. (2011). Assessing feigned cognitive impairment in defendants hospitalized for competency restoration: Further validation of the TOMI. *Journal of Forensic Psychology Practice*, *11*(4), 293–310. https://doi.org/10.1080/15228932.2011.562804

Colwell, K., Colwell, L. H., Perry, A. T., Wasieleski, D., & Billings, T. (2008). The test of malingered incompetence (TOMI): A forced-choice instrument for assessing cognitive malingering in competence to stand trial evaluations. *American Journal of Forensic Psychology*, *26*(3), 17–42. https://www.researchgate.net/publication/234115496_TOMI_Further_Validation_Final_Revision_012711

Cowden, V., & McKee, G. (1995). Competency to stand trial in juvenile delinquency proceedings: Cognitive maturity and the attorney-client relationship. *University of Louisville Journal of Family Law*, *33*, 629–659. https://heinonline.org/HOL/LandingPage?handle=hein.journals/branlaj33&div=43&id=&page=

Danzer, G. S., & Wasser, T. D. (2021, November). A literature analysis of the inventory of legal knowledge. *The Journal of the American Academy of Psychiatry and the Law*, *49*(1). Online. https://jaapl.org/content/early/2020/11/24/JAAPL.200042-20

Denney, R. L. (2008a). Evaluating competency to stand trial and sanity in the face of marked amnesia and claimed psychosis. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of Malingering Casebook* (pp. 413–427). Taylor & Francis.

Denney, R. L. (2008b). Negative response bias and malingering during neuropsychological assessment in criminal forensic settings. In R. L. Denney & J. P. Sullivan (Eds.), *Clinical* neuropsychology in the criminal forensic setting (pp. 91–134). Guilford Press.

Denney, R. L. (2012). Criminal forensic neuropsychology and assessment of competency. In G. J. Larrabee (Ed.), *Forensic neuropsychology: A scientific approach* (2nd ed., pp. 438–472). Oxford University Press.

Denney, R. L., & Fazio, R. L. (2021). Assessment of feigned cognitive impairment in criminal forensic neuropsychological settings. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (2nd ed., pp. 656–688). Guilford Press.

Everington, C. T., & Luckasson, R. (1992, 2022). *Competence assessment for standing trial for defendants with intellectual disability (CAST*ID): Test manual*. IDS Publishing Corporation.

Frederick, R. I., Crosby, R. D., & Wynkoop, T. F. (2000). Performance curve classification of invalid responding on the validity indicator profile. *Archives of Clinical Neuropsychology*, *15*(4), 281–300. https://doi.org/10.1093/arclin/15.4.281

Frederick, R. I., & Foster, H. G., Jr. (1991). Multiple measures of malingereing on a forced-choice test of cognitive ability. *Psychological Assessment*, *3*(4), 596–602. https://doi.org/10.1037/1040-3590.3.4.596

Gottfried, E., & Carbonell, J. (2014). The role of intelligence on performance on the inventory of legal knowledge (ILK). *The Journal of Forensic Psychiatry & Psychology*, *25*(4), 380–396. https://doi.org/10.1080/14789949.2014.920900

Green, P. (2005). *Green's word memory test for windows: User's manual* (revis ed.). Author.

Green, P. (2008). *Green's nonverbal medical symptom validity test: User's manual*. Green's Publishing.

Green, P., & Astner, K. (1995). *Oral word memory test: User's manual*. Cognisyst.

Green, P., Rohling, M., Lees-Haley, & Allen, L. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, *15*(12), 1045–1060. https://doi.org/10.1080/02699050110088254

Grisso, T., Steinberg, L., Woolard, J., Cauffman, E., Scott, E., Graham, S., Lexcen, F., Reppucci, N. D., & Schwartz, R. (2003). Juveniles' competence to stand trial: A comparison of adolescents' and adults' capacities as trial defendants. *Law and Human Behavior*, *27*(4), 333–363. https://doi.org/10.1023/A:1024065015717

Guenther, C. C., & Otto, R. K. (2010). Identifying persons feigning limitations in their competence to proceed in the legal process. *Behavioral Sciences and the Law*, *28*(5), 603–613. https://doi.org/10.1002/bsl.956

Hartmann, D. E. (2002). The unexamined lie is a lie worth fibbing: Neuropsychological malingering and the word memory test. *Archives of Clinical Neuropsychology*, *17*(7), 709–714. https://doi.org/10.1093/arclin/17.7.709

Hays, W. L. (1973). *Undefined* (2nd ed.). Holt, Rinehart, and Winston, Inc.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S., & Conference Participants. (2009). American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *23*, 1093–1129. https://doi.org/10.1080/13854040903155063

Heilbrun, K., Hawk, G., & Tate, D. C. (1996). Juvenile competence to stand trial: Research issues in practice. *Law and Human Behavior*, *20*(5), 573–578. https://doi.org/10.1007/BF01499043

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. Wiley-Interscience.

Hunter, S., Kois, L. E., Peck, A. T., Elbogen, E. B., & LaDuke, C. (2023). The prevalence of traumatic brain injury (TBI) among people impacted by the criminal legal system: An updated meta-analysis and subgroup analysis. *Law and Human Behavior*, *47*(5), 539–565. https://doi.org/10.1037/lhb0000543

Laboratory of Community Psychiatry, Harvard Medical School. (1973). *Competency to stand trial and mental illness*. Department of Health, Education, and Welfare. (DHEW Publication No. ADM77–103).

McGarry, A. L., Lelos, D., & Lipsitt, P. D. (1973). *Competency to stand trial and mental illness*. U.S. Government Printing Office.

McGuire, C., Crawford, S., & Evans, J. J. (2019). Effort testing in dementia assessment: A systematic review. *Archives of Clinical Neuropsychology*, *34*(1), 114–131. https://doi.org/10.1093/arclin/acy012

Nasreddine, Z. S., Phillips, N. A., B´edirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Of Directors, B., & Board of Directors. (2007). American academy of clinical neuropsychology (AACN) practice guidelines for neuropsychological assessment and consultation. *The Clinical Neuropsychologist*, *21*(2), 209–231. https://doi.org/10.1080/13825580601025932

Otto, R. K., Musick, J. E., & Sherrod, C. B. (2010). *Inventory of legal knowledge: Professional manual*. Psychological Assessment Resources.

Otto, R. K., Musick, J. E., & Sherrod, C. B. (2011). Convergent validity of a screening measure designed to identify defendants feigning knowledge deficits related to competence to stand trial. *Assessment*, *18*(1), 60–62. https://doi.org/10.1177/1073191110377162

Pankratz, L., & Erickson, R. C. (1990). Two views of malingering. *The Clinical Neuropsychologist*, *4*(4), 379–389. https://doi.org/10.1080/13854049008401832

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: Professional manual*. Psychological Assessment Resources.

Rey, A. (1958). L'examen clinique en psychologie *[The psychological examination]*. Presses Universitaries de France.

Riley, J. A. (1998, March). Introducing the revised-CAI and assessment of trial competency — use of the revised-competency assessment instrument handbook. Presentation at American Psychology–Law Society Biennial Conference, Redondo Beach, CA.

Roesch, R., Zapf, P., & Eaves, D. (2006). *FIT-R: Fitness interview test — revised: A structured interview for assessing competency to stand trial*. Professional Resource Press.

Rogers, R. (2007). Introduction. In R. Rogers (Ed.), *Handbook of diagnostic and structured interviewing* (pp. 1–19). Guilford Press.

Rogers, R. (2018). Detection strategies for malingering and defensiveness. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 18–41). Guilford.

Rogers, R., Robinson, E. V., & Henry, S. A. (2017). Feigned adjudicative incompetence: Testing effectiveness of the ILK and SAMA with jail detainees. *Assessment*, *24*(2), 173–182. https://doi.org/10.1177/1073191115605613

Rogers, R., Tillbrook, C. E., & Sewell, K. W. (2004). *Evaluation of competency to stand trial–revised*. Psychological Assessment Resources, Inc.

Rubenzer, S. (2011). Review of the inventory of legal knowledge. *Open Access Journal of Forensic Psychology*, *3*, 70–81. https://www.oajfp.com/_files/ugd/166e3f_17e7e50774cb4b2e8b625f1832af30e1.pdf

Sarker, S. P. (2011). In the twilight zone: Adolescent capacity in the criminal justices arena. *Advances in Psychiatric Treatment*, *17*(1), 5–11. https://doi.org/10.1192/apt.bp.108.006254

Savitsky, J., & Karras, D. (1984). Competency to stand trial among adolescents. *Adolescence*, *14*, 349–358. https://pubmed.ncbi.nlm.nih.gov/6464817/

Shipley, W., Gruber, C., Martin, T., & Klein, M. (2009). *Shipley-2 professional manual*. Western Psychological Services.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). McGraw Hill.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*(4), 545–561. https://doi.org/10.1076/1385-4046(199911)13:04;1-Y;FT545

Smith, T., Gildeh, N., & Holmes, C. (2007). The montreal cognitive assessment: Validity and utility in a memory clinic setting. *The Canadian Journal of Psychiatry*, *52*(5), 329–332. https://doi.org/10.1177/070674370705200508

Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. W., Schroeder, R. W., Suhr, J. A., & Conference Participants. (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 35, 1053–1106. https://doi.org/10.1080/13854046.2021.1896036

Tombaugh, T. N. (1996). *TOMM. test of memory malingering*. Multi-Health Systems.

Trzepacz, P. T., Hochstetler, H., Wang, S., Walker, B., & Saykin, A. J. (2015). Relationship between the montreal cognitive assessment and mini-mental state examination for assessment of mild cognitive impairment in older adults. *BMC Geriatrics*, *15*(1), 107. https://doi.org/10.1186/s12877-015-0103-3

Wahl, K. D., Vitacco, M. J., Panza, N. R., & Hyde, T. F. (2018). Utility of the inventory of legal knowledge in detecting feigning. *The Journal of Forensic Psychiatry & Psychology*, *29*(3), 483–497. https://doi.org/10.1080/14789949.2017.1421249

Watson, M. E., & Kivisto, A. J. (2017). The inventory of legal knowledge (ILK) and adults with intellectual disabilities. *Journal of Intellectual Disabilities and Offending Behaviour*, *8*(2), 83–89. https://doi.org/10.1108/JIDOB-10-2016-0018

Wildman, R., Batchelor, E., Thompson, L., Nelson, F., Moore, J., Patterson, M., & Deloasa, M. (1980). *The Georgia Court Competency Test: An Attempt to Develop a Rapid, Quantitative Measure for Fitness for Trial* [Unpublished manuscript]. Forensic Services Division, Central State Hospital.

Wygant, D. B., Sellbom, M., Ben-Porath, Y. S., Stafford, K. P., Freeman, D. B., & Heilbronner, R. L. (2007). The relation between symptom validity testing and MMPI-2 scores as a function of forensic evaluation context. *Archives of Clinical Neuropsychology*, *22*(4), 489–499. https://doi.org/10.1016/j.acn.2007.01.027

Wynkoop, T. F., & Denney, R. L. (1999). Exaggeration of neuropsychological deficit in competency to stand trial. *Journal of Forensic Neuropsychology*, *1*(2), 29–53. https://doi.org/10.1300/J151v01n02_04

Wynkoop, T. F., & Denney, R. L. (2005). Test review: Green's word memory test (WMT) for windows. *Journal of Forensic Neuropsychology*, *4*(3), 101–105.