

# A Literature Analysis of the Inventory of Legal Knowledge

Graham S. Danzer, PsyD, and Tobias D. Wasser, MD

The Inventory of Legal Knowledge (ILK) is a feigning measure of growing usage, familiarity, and controversy in research and practice. A comprehensive review of a smaller literature base yields recurring themes in the ILK literature. There were mentions of feigned lack of legal knowledge tending to associate with feigned psychopathology, concerns about false positives at or around the ILK cutoff score, and potential complications when the ILK is administered to individuals with very low intellectual functioning. Possible underestimation of both false positives and false negatives suggests there may be a need for a revised edition of the measure, further item discrimination, and a meta-analysis of extant research studies with similar designs. An ILK-2 with required user training, as well as a lower cutoff score, an indeterminate range, and weighting of scores in association with criminal history may increase the measure's validity, robustness, and utility in larger assessments of malingering.

**J Am Acad Psychiatry Law 49(1) online, 2021. DOI:10.29158/JAAPL.200042-20**

**Key words:** Inventory of Legal Knowledge; ILK; feigning; malingering; forensic; competence to stand trial

Ensuring defendants are competent to stand trial helps ensure the fairness, integrity, and indeed the justice of the criminal justice system.<sup>1-3</sup> Accordingly, standards for raising questions about defendants' competence are fairly low.<sup>4</sup> It has been estimated in research articles that about 60,000 defendants in the United States may have their competence questioned annually.<sup>5,6</sup>

Malingering trial incompetence is substantially incentivized in higher-stakes criminal justice proceedings, may challenge the validity of psychological assessments, and is often among the major determinants of outcome in forensic referrals.<sup>7-9</sup> Successful malingering may result in delays or evasion of prosecution, maintain the relative comfort of hospitalization as opposed to jail, lead to the acquisition of controlled medications, and encourage the court to consider diversionary measures or mitigation.<sup>2,4,8,10</sup> Thus, it is unsurprising that malingering is not generally considered rare, with researchers and experts estimating it occurs in 10 to 50 percent of

competency cases (with an estimate of 20 percent mentioned more frequently) and in an estimated 37 percent of murder cases.<sup>1,2,4-6,8-12</sup> In other words, when criminal charges and possible penalties are more serious, the motivation to malingering is proportionally enhanced.

For these reasons, it is necessary in competency evaluations to assess response style (i.e., candid, defensive, irrelevant, feigned, malingered) routinely and systematically via current, objective, structured, research-supported, and highly reliable and valid measures.<sup>1,3,4,7-10,12,13</sup> Incorrectly labeling genuine responders as feigning or malingering may result in individuals being returned to court despite being too mentally ill to participate meaningfully in their defense.<sup>1,2</sup> In such cases, the criminal proceedings and outcomes would be fundamentally unfair. Conversely, failing to identify accurately individuals feigning or malingering may further burden an under-resourced mental health system, delay access to treatment for some defendants, facilitate unnecessary treatment for others, and obstruct justice.<sup>1,2</sup> Failing to identify malingering accurately may also result in particularly criminogenic and dangerous individuals being hospitalized alongside highly vulnerable, severely compromised peers. In hospitals, such individuals are necessarily afforded patients' rights generally

---

Published online November 24, 2020.

Dr. Danzer is a Senior Psychologist at Florida State Hospital, Chattahoochee, FL. Dr. Wasser is the Chief Medical Officer at Whiting Forensic Hospital, Middletown, CT, and Assistant Professor of Psychiatry at Yale School of Medicine, New Haven, CT. Address correspondence to Dr. Danzer at: gdanzer@alliant.edu.

Disclosures of financial or other potential conflicts of interest: None.

prohibiting coercive interventions in all but the most imminently and obviously dangerous scenarios.

Accordingly, researchers advise a lower threshold in screening for suspected feigning (i.e., making many errors on other tests, claiming amnesia of events surrounding arrest, etc.).<sup>4,10,14</sup> Balancing the low threshold for suspicion, researchers consistently recommend maintaining a higher standard for concluding malingering based on conservative decision-making rules and cross-validation strategies (i.e., use of multiple data sources and tests).<sup>15,16</sup> Notably, cross-validation strategies have been found in research studies to associate with about 90 percent accuracy of conclusions.<sup>14</sup>

While defendants may also feign psychopathology or cognitive impairment, the unique nature and substantial incentives of competency proceedings (particularly in cases of serious crimes) makes it essential for assessments of response style to rule out the possibility of feigned lack of legal knowledge.<sup>1,2,4-6,8,12</sup> Low levels of legal knowledge do not necessarily presume incompetence, because deficits can often be remedied through basic legal education and clarification from defense counsel. At least a basic understanding of the legal system is necessary to be found competent, however, which may therefore incentivize feigning knowledge deficits.<sup>5,7</sup> Hence, it is not uncommon for defendants to report ignorance of legal concepts or give outlandish answers when asked questions about basic courtroom or criminal procedures.<sup>4,12</sup>

We conducted a literature review on the validity, usage, and applications of the Inventory of Legal Knowledge (ILK).<sup>8,17</sup> This effort follows Scott and McDermott's<sup>13</sup> identification of the measure's relevance in forensic training, its growing use particularly in higher stakes criminal cases, and the emerging emphasis on utilizing objective tests in the competence evaluation and restoration process. This paper summarizes the entirety of available and acquired ILK research, presents implications for its use by forensic practitioners, and offers suggestions for a possible future revision of the measure.

### The Inventory of Legal Knowledge

Developed by Musick and Otto, the ILK is a measure of effort to feign or exaggerate a lack of legal knowledge relevant to adjudicative competency.<sup>4,9,17,18</sup> It was among the first psychological

tests developed specifically to assess response style in competency proceedings.<sup>10,19,20</sup> It was innovative in its focus on trial-related concepts such as the role and functions of defense counsel and other key legal officials.<sup>1,3</sup>

Importantly, the ILK is not a psychological assessment of a defendant's ability to understand and participate intelligently in the legal process, as high or low scores do not presume a stronger or weaker knowledge base.<sup>5,6,18</sup> Rather, it is often useful in cases where a defendant is suspected of feigning or exaggerating a lack of legal knowledge.<sup>18</sup> In such cases, ILK performance should be considered alongside other sources of information as part of a multi-data source evaluation of response style.<sup>6,7,18</sup>

### Administration and Interpretation of the ILK

The ILK is a two-alternative, forced-choice measure with 61 true-false items focused on various aspects of the criminal-legal system and its operation.<sup>6-8,20</sup> The interviewed defendant is read each item aloud, instructed to answer each question true or false, and receives immediate feedback on whether or not each answer was correct.<sup>5,6,9,16</sup> The ILK usually takes about 15 minutes to administer, one minute to score, and costs slightly over \$2 per administration.<sup>16</sup> Theoretically, immediate feedback should amplify feigning and encourage honest though incorrect responders to refocus and answer questions more accurately.<sup>8</sup> The ILK is intended for individuals age 12 years or older who can read and comprehend English at a fifth-grade level or higher and have at least a basic familiarity with the U.S. legal system.<sup>4,8,16,20</sup>

Each of the 61 items answered correctly is scored one point, for a total score ranging from 0 to 61.<sup>8</sup> The standardized cutoff score ( $\leq 47$ ) was found to optimally balance concerns regarding sensitivity (i.e., the proportion of those who were feigning and correctly identified as such by the test) and specificity (i.e., accurate classification of non-feigners).<sup>10,20</sup> Total scores between 24 and 36 (i.e.,  $30 \pm 6$ ,  $P < .05$ ) are in the range obtainable by chance; scores below 24 exceed the 95 percent confidence interval and were reportedly obtained by 50 percent of feigning defendants in studies noted in the manual; hence, these scores should be interpreted as providing compelling evidence of a feigned response style.<sup>4-9,16,20</sup> Poor performance can also be due to random

responding.<sup>7,16</sup> Nevertheless, lower total ILK scores make feigning a more reasonable conclusion.<sup>5,7,16</sup>

ILK scores are interpreted primarily via significantly below chance performance (i.e., symptom validity testing) and floor strategies.<sup>6,9</sup> Long the gold standard in malingering assessment, below chance performance refers to the lowest score that could reasonably be expected of examinees who were purely guessing.<sup>2,8,16</sup> Thus, scores further below chance more conclusively suggest the individual knows the correct answer and chooses not to offer it. Specific to the ILK, below chance scores have high specificity (virtually no genuine responders will score this low), but low sensitivity, because such a poor performance occurred in less than 1 percent of clinical and forensic samples.<sup>9,16</sup> In different ILK studies, there were indications of below chance criteria associating with zero subjects classified incorrectly as feigning (i.e., false positives), though 18 to 43 percent of participants were classified incorrectly as not feigning (i.e., false negatives).<sup>4,7,9,16</sup>

Floor strategies involve the administration of such simple tasks that even very cognitively impaired individuals should be able to complete them (i.e., improbable failure).<sup>8,9</sup> Under the floor-effect methodology, an individual's score is compared with the average scores obtained by genuine responders in a comparable or relevant reference group and considered suggestive of feigning when the score is significantly below the comparison group average.<sup>6,7</sup> Floor-effect methodologies have better sensitivity than below-chance methods.<sup>7</sup> The ILK's assumed level of improbable failure has not been sufficiently empirically validated.<sup>9</sup> Accordingly, floor-effect strategies can be problematic when applied to severely impaired populations.<sup>9</sup> To balance sensitivity and specificity, it is important to supplement the below-chance methodology with a test of the floor effect.<sup>8</sup>

### Test Development

As summarized by Rubenzer,<sup>16</sup> the initial development of the ILK involved the creation of over 100 items grouped into 11 categories. Items considered trivial, confusing, overly technical, or redundant were eliminated. Retained items were further revised for clarity, determined via the Flesch-Kincaid grade-level formula to require at least a fifth-grade reading and comprehension level, and reviewed by a panel of nine forensic experts and lawyers

for problems related to item format and bias (i.e., related to race, gender). Items were further revised before and after the ILK was first administered to research participants.

From 2010 to 2011, Otto and colleagues<sup>6,20</sup> conducted foundational ILK studies on five samples of adults; these studies are often referenced in other studies, unpublished dissertations, and perspective papers.<sup>3-5,7-10,16,18</sup> There were initial samples of 100 community-based psychiatric inpatients, 207 college students, 110 defendants in competency proceedings, and 99 insanity acquittees. All subjects were administered the ILK and randomly assigned to groups instructed to respond honestly or fake bad (intentionally perform poorly).

Major results from foundational studies were as follows. There were significant differences in ILK scores of honestly responding psychiatric patients ( $M = 53.04$ ,  $SD = 7.49$ ), genuinely incompetent defendants ( $M = 40.6$ ,  $SD = 8.53$ ), and college students ( $M = 55.37$ ,  $SD = 3.52$ ). Founding ILK studies also yielded large to very large Cohen's  $d$  differentiation of malingerers (1.62) and insanity acquittees and nonmalingerers (1.43), as well as good internal consistency (Cronbach's  $\alpha = 0.88$ ) and two-week test-retest reliability ( $r = 0.76$ ). The standard error of measurement (SEM) ranged from 2.05 to 2.77 points (plus or minus). The 95 percent confidence interval was  $\pm 5.5$  points. These results were presented by Otto and colleagues<sup>6,20</sup> as evidence in support of the ILK as a reasonably valid measure of effort and differentiating honest responders from feigners.

### Strengths and Limitations of the ILK

#### Strengths

The ILK has been identified as one of the most important feigning measures for trainees and residents to learn.<sup>13</sup> It is relatively brief, simple, and straightforward, presents as a true-false assessment of legal knowledge rather than a more obvious feigning test, incorporates prompts and cues to clarify the underlying response style, and is more directly relevant to competency than other cognitively-focused feigning measures, such as the Test of Memory Malingering (TOMM).<sup>6,8,9,12,18</sup> Because items are read aloud, the ILK can also be used with illiterate defendants capable of processing English language at a fifth-grade level or higher.<sup>8</sup>

### Limitations

The ILK normative samples were relatively small, not usually in competency proceedings (i.e., with less motivation to feign convincingly), overrepresented in terms of female defendants, and had known groups performing poorly (i.e., possible criterion contamination).<sup>5-7,16</sup> Furthermore, not all ILK items call for common legal knowledge, some items are worded in confusing double negatives, repeated administrations of the measure may have practice effects, and scores may be affected by severe psychopathology, limited education, and random responding.<sup>2,6,7,9,20</sup>

Another area of recurring researcher concern is that the current ILK cutoff score of  $\leq 47$  appears to be associated with higher levels of false positives (e.g., 0.21 for a clinical sample noted in the ILK manual).<sup>1,3,5,9-11,16</sup> This may create complications when assessing the response style of defendants who are very impaired and questionably motivated.<sup>5</sup> Reducing the cutoff score (e.g., below chance, or  $< 24$ ) would greatly decrease false positives, but it would substantially increase false negatives as well, and thus permit more feigning individuals to escape detection.<sup>2,5,6</sup> The resulting need to balance sensitivity and specificity was the basis for Otto and colleagues deciding on a cutoff score of  $\leq 47$ , despite larger numbers of genuinely incompetent defendants in the validation sample scoring well-below this threshold ( $M = 40.6$ ,  $SD = 8.53$ ).

Of note, the ILK validation sample did not include a group of individuals with intellectual and developmental disabilities (IDD).<sup>1,6,16</sup> This is concerning because 4 to 10 percent of all criminal defendants were reported to have some extent of intellectual deficits<sup>3,11</sup>; in one research study, 38 percent of 1,710 federal defendants had some form of neurologic impairment.<sup>21</sup> Impaired cognitive functioning may significantly complicate performance on psychological tests, resulting in genuine performance appearing similar to deliberate feigning, thereby increasing the likelihood of misclassification (i.e., false positives).<sup>3,11</sup> Thus, the ILK may not be valid with IDD populations, particularly those who are more severely impaired.<sup>6,12,16,18</sup>

### Methods

We searched for ILK research in major scholarly databases, including Psych Info, Psych Tests, Psych Articles, Psychiatry Online, Science Direct, Google

Scholar, PubMed, and *The Journal of the American Academy of Psychiatry and the Law*. This search yielded a relatively smaller base of published studies. Specifically, there were no literature reviews since Rubenzer<sup>16</sup> in 2011, although we identified 14 subsequent peer-reviewed journal articles, including seven empirical studies, three unpublished dissertations, and two references in major (if not authoritative) textbooks.<sup>15,22</sup> Recent publications more deeply explored or addressed limitations referenced in the initial ILK manual, cited in follow up research studies, and frequently recalled and contested in forensic practice.

To compensate for this relative paucity of literature, we reviewed and incorporated all available studies on the ILK, as well as practice guideline papers, commentaries in major texts, and unpublished dissertations. The reviewed studies are presented in Table 1. This review emphasizes sources that are empirical, recently published (i.e., within the last five years), and more frequently cross-referenced in other publications. Given the practice component of this article, it was essential not only to identify recurring results and findings, but also concurring researcher opinions.

### Results

Our review yielded four simulation studies of college students and jail detainees, with samples ranging from 70 to 195 participants.<sup>2,5,9,10</sup> Three studies involved exclusively real defendants in competency proceedings, two of which were in state hospital facilities (the third did not specify the setting), with sample sizes ranging from 120 to 203.<sup>1,5,8</sup> Finally, there was one study involving 65 college students and 65 defendants in competency proceedings in a state hospital.<sup>11</sup>

Recent studies largely affirmed and supplemented earlier foundational studies' encouragements for future research to explore different ILK cutoff scores, validity with different populations, convergence with other response-style measures, and with multi-site sample sizes large enough (i.e., at least 150 subjects) to suggest results were not likely attributable to error.<sup>6,8,16</sup> Of note, only two of eight empirical ILK studies had samples with at least 150 subjects, both of which were conducted by Gottfried and colleagues.<sup>1,5</sup> These studies indicated that the standard cutoff score of  $\leq 47$  was associated with lower numbers of false positives (i.e., three percent among

**Table 1** ILK Studies and Implications

Empirical Study Sample	Validity-Related Results	Implications for Future ILK Revision
120 competency examinees, setting unspecified <sup>8</sup>	ILK correlations with Rey FIT = 0.67, RDS = 0.63, and TOMM-Trial 2 = 0.60 (all $P < .01$ )	Significant convergence with other measures of effort or cognition support use of the ILK as a measure of response style
70 jail detainees <sup>9</sup>	ILK cutoff score of $\leq 47$ ; PPP = 0.60 at 15% base rate, PPP = 0.74 at 25% base rate	Positive preliminary findings using well-defined and revised cutoff scores, as well as the R-ILK-90 and R-ILK-95 sub-measures
195 college students <sup>5</sup>	ILK/M-FAST correlation = $-0.51$ ( $P < .01$ ); ILK score $\leq 47$ had PPP = 0.92, NPP = 0.60, sensitivity = .35, specificity = 0.97	Revised ILK cutoff score $\leq 40, 42$ may balance concerns about false positives and negatives
130 state hospital patients in competency proceedings <sup>5</sup>	ILK/M-FAST correlation = $-0.54$ ( $P < .01$ ); ILK score $\leq 47$ had PPP = 0.80, NPP = 0.67, sensitivity = 0.57, specificity = 0.86	ILK cutoff score from $\leq 35$ to $\leq 42$ may better balance concerns about false positives and negatives
203 state hospital patients in competency proceedings <sup>1</sup>	ILK/M-FAST correlation = $-0.49$ ( $P < .01$ ); higher cutoff score (ILK $\leq 35$ , M-FAST $\geq 16$ ), correlation ( $\chi^2 = 14.72, P < .001$ )	Feigning individuals often use more than one strategy, particularly in cases of very low ILK scores or very high M-FAST scores
100 college students <sup>10</sup>	ILK/M-FAST correlation = $-0.55$ ( $P < .001$ , Cohen's $d = 1.48$ ); ILK and CAST-MR correlations ranged from 0.80 to 0.90	Feigning individuals often use more than one strategy; balance of sensitivity and specificity not highly improved at ILK cutoff score $\leq 37$ , or $\leq 42$
65 state hospital patients in competency proceedings, 65 college students <sup>11</sup>	ILK/M-FAST correlation = $-0.52$ , ( $P < .01$ ); Shipley-2 intelligence/ILK score correlation = 0.39 ( $P = .002$ )	Affirms manual cautions on ILK scores and IDD; further study needed on item discrimination and severely impaired groups
32 adults with IDD, most had no active legal cases <sup>3</sup>	ILK/M-FAST correlation = $-0.34$ (moderate effect size); ILK and IQ (KBIT-2) correlation was significant ( $P < .05$ )	Affirms manual cautions regarding ILK scores and IDD; nonsignificant convergent validity with TOMM
Dissertations: Each had a sample of 40–73 <sup>2,4,7</sup>	Manual-comparable rates of sensitivity, specificity, and reported false positives	As before, affirmed manual cautions about ILK scores and $<$ fifth-grade reading level, acculturation, effort, and lower cognition or IQ

ILK, Inventory of Legal Knowledge; FIT, Rey 15-Item Test; RDS, Reliable Digit Span; TOMM, Test of Memory Malingering–Trial 2; M-FAST, Miller Forensic Assessment of Symptoms Test; PPP, positive predictive power; NPP, negative predictive power; CAST-MR, Competence Assessment for Standing Trial for Defendants with Mental Retardation; IDD, intellectual and developmental disabilities.

college student simulators, 14 percent among real defendants in competency proceedings) than what was suggested elsewhere in the literature.<sup>1,3,5,9–11</sup>

As referenced in Table 1, one study that included 120 real defendants in competency proceedings indicated moderate positive correlations between ILK total scores and scores on other tests of effort and feigning cognitive or memory impairment, including the Rey 15 Item Test (FIT), Reliable Digit Span (RDS), and the TOMM Trial 2.<sup>8</sup> Five studies compared total scores on the ILK and the Miller Forensic Assessment of Symptoms Test (M-FAST).<sup>1,3,5,10,11</sup> Of those ILK/M-FAST studies, two included college student samples of 100 and 195 subjects,<sup>5,10</sup> two involved 130 and 203 defendants in competency proceedings in hospitals,<sup>1,5</sup> one involved 65 college students and 65 hospitalized defendants in competency proceedings,<sup>11</sup> and one involved a sample of 32 adults with a diagnosis of IDD (mean IQ = 60, i.e., more significant impairment).<sup>3</sup> All five ILK/M-FAST studies (with samples between 100 and 205) reported M-FAST/ILK ranging from  $-0.49$  to  $-0.55$

( $.001 < P < .01$ ).<sup>1,5,10,11</sup> The smaller study of 32 adults with IDD reported a weak to moderate though still significant M-FAST/ILK correlation of  $-0.34$ , as well as an M-FAST/TOMM correlation reaching the standard level of significance ( $P \leq .05$ ).<sup>3</sup>

Studies often raised concerns about the ILK's predictive and discriminant validity. Researchers in seven of the eight empirical studies referenced higher numbers of false positives when defendants scored at or near the cutoff score.<sup>1,3,5,9–11</sup> These points were at times the basis for researchers recommending that a score  $\leq 47$  may more conservatively raise suspicion of feigning and suggest the need for a comprehensive, multi-sourced evaluation.<sup>1,10</sup>

As noted by Rogers et al<sup>9</sup> and in dissertations,<sup>4,7</sup> there were recurring questions about the discriminant validity of ILK items 20, 25, 43, and 44. Gottfried and Carbonell's<sup>11</sup> hospital subsample of 65 subjects yielded weak to moderately positive correlations, both for IQ and ILK total scores ( $r = 0.38, P < .01$ ) and for IQ and ILK items 12, 34, 20, 5, and 50 (arranged in order starting with the strongest

correlate;  $0.25 \leq r \leq -0.44$ ,  $P \leq .05$ ). Items 12 and 34 were the most strongly correlated with IQ ( $P < .01$ ) and were both related to testimony.<sup>11</sup> Of the four ILK items recurrently identified as of concern to Rogers<sup>9</sup> and in dissertations,<sup>4,7</sup> only item 20 was also significantly correlated with IQ scores in the study by Gottfried and Carbonell.<sup>11</sup> For the smaller study of 32 adults with IDD,  $M \pm SD$  ILK scores were  $33.96 \pm 4.69$ , although mean IQ scores, severity of impairment, and levels of reading or English comprehension were not reported.<sup>3</sup>

### Complications in Practice

The ILK was developed as a standalone psychological test of feigned lack of legal knowledge and was not intended as a substitute for a comprehensive evaluation of response style.<sup>6</sup> There are concerns that the ILK and other brief measures may at times be used unilaterally to indicate feigning and malingering (i.e., as an alternative to a more comprehensive evaluation).<sup>12,15</sup> Inadequate evaluations of response styles may occur due to limited evaluator training or motivation, or perhaps due to concerns about prejudicial versus probative evidence.<sup>12,23</sup> Nevertheless, local evaluator assessments and recommendations are overwhelmingly relied upon by the judicial system, regardless of common perceptions of competency evaluations as low quality and pro-defense.<sup>15</sup> In part, this may be due to long-standing evaluator-judiciary relationships, deferring difficult decision-making to experts, or because judges overriding evaluator opinions could become grounds for due process appeals.<sup>15,24</sup>

The ILK's use in practice may also be complicated by malingering coaching, practice effects of repeated administrations, and inadequate evaluator consideration of presumable legal knowledge and criminal sophistication.<sup>4,7,11,12,20</sup> For these reasons, prior assessments of genuine impairment, obviously higher levels of impairment, and scoring above the ILK cutoff score may not necessarily presume a candid response style or a lack of motivation to evade prosecution, particularly for serious crimes (or when genuine responding may not necessarily be rewarded).<sup>12,16</sup> For example, it is not uncommon for obviously severely impaired defendants to be found incompetent, admitted to hospitals, receive treatment, stabilize, become more rationally appreciative of alleged evidence against them, consider possible consequences of conviction, and then take purposeful steps to remain hospitalized.

These possibilities reasonably call into question a core assumption in many ILK research studies. In research, judicial determinations of incompetency are often interpreted as presuming defendants were not feigning or malingering. Psychological test findings contrary to judicial decisions were often presumed to be indicative of false positives. While not discrediting researcher concerns about false positives and potentially serious ramifications, there may also be a connection between influences on judicial decision-making and less frequent researcher concerns about possible underestimation of false negatives.<sup>1,5,12</sup>

### Future Research

Our review highlights a need to clarify the limitations of the current ILK cutoff score.<sup>9,16</sup> Rogers and colleagues<sup>9</sup> proposed that a future revision of the measure should have an indeterminate range one standard deviation above and below the mean (i.e., 31 to 49 for incompetent defendants). Consistent with this position, the results of the simulation study by Wahl and colleagues<sup>10</sup> suggested that ILK scores  $\leq 47$  may more conservatively raise concern about response style, thereby warranting a comprehensive, multi-sourced evaluation. While acknowledging differing effects on sensitivity and specificity, researchers consistently indicated possible benefits of lowering the cutoff to somewhere between 28 and 42.<sup>1,2,5,6,16</sup>

Rubenzler's review on the ILK<sup>16</sup> offered important considerations related to the possibility of a lowered cutoff score. The lowest score in the manual for which full statistics were reported was  $\leq 36$ , which still indicated a false-positive rate of 0.08. This false-positive rate is problematic because 0.08 is beyond the generally accepted level of error in psychological testing ( $P \leq .05$ ). Rubenzler<sup>16</sup> noted ILK scores  $\leq 33$  are increasingly suggestive of feigning, whereas 28 was the highest score at which false positives appeared to occur  $< 5$  percent of the time. Rubenzler also noted that the score of 28 would also be approximately one standard deviation below the mean for the smaller sample of 32 adults with IDD.<sup>3</sup> In other words, there may be future research questions about how a revised ILK cutoff score of 28 might affect the sensitivity-specificity balance, and whether it may be interpretable for IDD populations (of lesser impairment) by using a floor strategy or theory of improbable failure.<sup>3</sup>

Taking the points of researchers a step further, the ILK's validity and reliability may be further

improved if a revised measure somehow incorporated prior criminal history into score weighting and interpretation.<sup>7,16</sup> Although a connection between antisocial personality features and feigning or malingering cannot be presumed, criminal history is relevant to evaluations of competence and response style because it can in some cases be admissible as character or propensity evidence (i.e., a higher likelihood of being found guilty), may aggravate sentencing, and may thereby enhance motivation to malingering. Further, a longer standing and more sophisticated criminal history may reasonably presume a greater extent of legal knowledge and familiarity with the criminal justice system. In cases of significant or sophisticated criminal history, however, it would also be important to clinically determine whether poorer-than-expected performance (i.e., floor effect or improbable failure) may be attributable to less strategic forms of oppositionality. The need to evaluate this possibility is consistent with researcher recommendations to always consider alternative hypotheses for poor test performance.<sup>7,16</sup>

To better address these limitations, future research should focus primarily on clinical rather than college student samples.<sup>10</sup> Updated norming should include larger groups, as well as subsamples of adolescents and young adults, individuals of lower and higher intellectual functioning, and individuals of lower and higher psychotic symptom-related impairment.<sup>3,6,8,11,16</sup> Factor analytic studies may determine the presence of a single ILK construct or discreet loadings.<sup>16</sup> Empirical studies should guide ILK-2 development, in part by identifying and removing less discriminant items.<sup>4,7,9,11</sup> The higher number of prior studies using the ILK and M-FAST offers a potential database for conducting meta-analytic or larger-scale item-discrimination studies.

Given concerns about false positives, future norming should include an additional screen or embedded measure of effort.<sup>9,11,16</sup> Options include Reliable Digit Span (RDS), the Rey-FIT, and the TOMM, which were noted in earlier research to associate moderately positively with ILK scores.<sup>3,8</sup> Dissertations have used the Rey-FIT and the TOMM for similar purposes.<sup>2,10</sup> Not all studies in this review, however, reported significant convergence between ILK and TOMM scores.<sup>3,11</sup>

In addition, future research should seek to determine whether possible false positives in cases of reportedly lower intellectual functioning may be attributable to feigned cognitive impairment and feigned lack of legal knowledge. This possibility was

not expressively considered in most low intellectual functioning studies and critiques, despite previously discussed research findings (i.e., ILK and M-FAST studies) consistently indicating a tendency for feigning individuals to often utilize more than one strategy.<sup>1,3,5,8-11</sup> Future research should also explore a possible association between underestimated false negatives and secondary influences on judicial decision-making, effects of criminal history and sophistication, and test validity complications associated with volitional effort testing only recently gaining wider spread appreciation and familiarity among forensic practitioners.<sup>4,7,9,11,15,22</sup>

These limitations suggest at least the possibility that dubiously low ILK scores may reasonably call earlier IQ scores into question. This possibility is of particular consideration in cases where criminal history suggests a higher level of adaptive functioning and psychological sophistication, and the individual's objective history does not support alternative hypotheses (i.e., random responding, character-driven forms of oppositionality). More objective data on false positives and false negatives may be acquired if future studies used an outcome measure other than judicial decisions and better accounted for family-wise error by running fewer tests.

## Conclusions

The ILK has been advanced by Otto and colleagues as a psychological test of feigning a lack of legal knowledge. It was originally developed with the goal of optimally balancing concerns about sensitivity and specificity. A fairly comprehensive review of the available literature base indicates generally favorable though nonconclusive research findings. This review also suggests clinical utility of the measure, particularly in higher-stakes criminal cases, when the motivation to malingering is presumably higher.

Of the eight empirical studies, seven raised concerns about high rates of false positives. Two of those studies indicated possible complications with IDD populations. False positives were of higher concern in cases where defendants were clearly or more severely impaired and scored at or near the standard ILK cutoff score ( $\leq 47$ ). False positives may have been overestimated due to practice-based factors reminiscent of criterion contamination. Further, seven of the eight empirical studies had small sample sizes or included college student populations with

less incentive (and perhaps less criminal wherewithal) to feign convincingly.

A possibility not expressly considered in the reviewed studies is that poorer ILK scores may call the validity of older IQ scores into question, as volitional effort testing has only recently been advanced. This possibility suggests that IQ may have appeared low because the possibility of low effort was not better accounted for. This alternative hypothesis would be supported in cases where the individual's criminal history suggests a level of sophistication and adaptive functioning inconsistent with IDD.

Similarly, studies critiquing the ILK frequently included concerns about a lack of norming on IDD populations, though did not appear to consider possible multicollinearity between IDD and reading-level considerations, as Rubenzer<sup>16</sup> noted about the ILK manual and other preliminary graduate research. Whereas the potentially serious ramifications of false positives cannot be overstated, these latter points are consistent with other publications indicating the possibility of underestimating false negatives.<sup>2,5,6</sup>

Research findings, scholarly positions, possible overestimation of false positives and false negatives, and emerging alternative hypotheses collectively support the need for a future revision of the ILK measure. Extant research suggests classification errors may be reduced by lowering the ILK cutoff score or replacing it with an indeterminate range plus or minus one standard deviation from the mean (i.e., a well-defined cutoff score). Improved item discrimination may be facilitated by removing the four items discussed in this article as potentially problematic. As identified by Rogers and colleagues,<sup>9</sup> replacement items may be guided or in some ways patterned on the 40 ILK items answered correctly by  $\geq 90$  percent of research participants (i.e., the R-ILK-90) and the 28 items answered correctly by  $\geq 95$  percent of participants (i.e., the R-ILK-95).

Additionally, the ILK may be improved if it included secondary scales attempting to identify random responding, oppositionality, and response style coaching. The ILK may be further improved if items were weighted to account for presumable legal knowledge (i.e., ability to answer forced-choice items correctly) in light of criminal history. An alternative criterion to judicial determination of competency appears warranted. ILK developers might consider requiring test administrators to complete a standardized training and perhaps annual recertification.

Applied to practice, these modifications may help reduce false positives and false negatives, thereby enhancing the ILK's contribution to comprehensive, valid, and reliable assessments of malingering incompetency.

## References

1. Gottfried E, Vitacco M, Steadham J: The relationship between feigned psychiatric symptoms and feigned lack of legal knowledge among a multi-site sample of forensic inpatients. *Int J Forensic Ment Health* 16:197–206, 2017
2. McLaughlin J: Using the ILK with foreign-born adults in the United States. Unpublished doctoral dissertation, Sam Houston University, 2016
3. Watson M, Kivisto A: The Inventory of Legal Knowledge (ILK) and adults with intellectual disabilities. *Jnl Intell Disabil Offen Behav* 8:83–9, 2017
4. Gaskell M: The effect of neuropsychological impairment feigned adjudicative incompetence on the Inventory of Legal Knowledge. Unpublished doctoral dissertation, Xavier University, 2016
5. Gottfried E, Hudson B, Vitacco M, Carbonell J: Improving the detection of feigned knowledge deficits in defendants adjudicated incompetent to stand trial. *Assessment* 24:232–3, 2017
6. Guenther C, Otto R: Identifying persons feigning limitations in their competence to proceed in the legal process. *Behav Sci & L* 28:603–13, 2010
7. Campbell K: Inventory of legal knowledge: an examination of psychometric properties in an inpatient setting. Unpublished master thesis, James Madison University, 2017
8. Otto R, Musick J, Sherrod C: ILK: Inventory of Legal Knowledge Professional Manual. Lutz, FL: Psychological Assessment Resources, 2010
9. Rogers R, Robinson E, Henry S: Feigned adjudicative incompetence: testing effectiveness of the ILK and SAMA with jail detainees. *Assessment* 24:173–82, 2017
10. Wahl K, Vitacco M, Panza N, Hyde T: Utility of the Inventory of Legal Knowledge in detecting feigning. *J Forensic Psychi Ps* 29:483–97, 2018
11. Gottfried E, Carbonell J: The role of intelligence on performance on the Inventory of Legal Knowledge (ILK). *J Forensic Psychi Ps* 25:380–96, 2014
12. Rubenzer S: How competency examiners should (and often don't) assess for malingering and poor effort. *Prosecutor* 50:1–22, 2017
13. Scott C, McDermott B: Teaching psychological assessment to forensic psychiatry fellows: a practical guide. *Acad Psychiatry* 41:803–6, 2017
14. *Clinical Assessment of Malingering and Deception* (3rd edition). Edited by Rogers R. New York: Guilford Press, 2008
15. Melton G, Petrila J, Poythress N, *et al*: *Psychological Evaluations for the Courts: a Handbook for Mental Health Professionals and Lawyers*. New York: Guilford Press, 2017
16. Rubenzer S: Review of the Inventory of Legal Knowledge. *Open Access J Forensic Psychol* 3:70–81, 2011
17. Musick J, Otto R: *Inventory of Legal Knowledge*. Lutz, FL: Psychological Assessment Resources, 2006
18. Acklin M: The forensic clinician's toolbox. I: a review of competency to stand trial (CST) instruments. *J Pers Assess* 94:220–2, 2012
19. Fogel M, Schiffman W, Mumley D, *et al*: Ten year research update (2001–2010): evaluations for competence to stand trial (adjudicative competence). *Behav Sci & L* 31:165–91, 2013
20. Otto R, Musick J, Sherrod C: Convergent validity of a screening measure designed to identify defendants feigning knowledge



## Danzer and Wasser

- deficits related to competence to stand trial. *Assessment* 18:60–2, 2011
21. Cochrane R, Grisso T, Frederick R: The relationship between criminal charges, diagnoses, and psycholegal opinions among federal pretrial defendants. *Behav Sci & L* 19:565–82, 2001
  22. Rogers R, Bender S: *Clinical Assessment of Malingering and Deception* (4th Edition). New York: Guilford Press, 2018
  23. Fitch W: AAPL practice guideline for the forensic psychiatric evaluation of competence to stand trial: an American legal perspective. *J Am Acad Psychiatry Law* 35:509–13, 2007
  24. Martinez M: Application of a systematic model for ethical decision making in forensic psychology practice. Presented at the Annual Convention for the American Psychological Association, Chicago, IL, August 2019