Construction of a Forced-Choice Task for the Assessment of Factual Understanding and Feigning in Competency to Stand Trial Evaluations

by

Elizabeth Arias

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2011

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date	Chair of Examining Committee Michele Galietta, Ph.D.
Date	Executive Officer Maureen O'Connor, JD, Ph.D.

Keith Markus, Ph.D.

Diana Falkenbach, Ph.D.

Nancy Ryba, Ph.D.

Phillip Erdberg, Ph.D., ABPP

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

CONSTRUCTION OF A FORCED-CHOICE TASK FOR THE ASSESSMENT OF FACTUAL UNDERSTANDING AND FEIGNING IN COMPETENCY TO STAND TRIAL EVALUATIONS

by

Elizabeth Arias

Advisor: Professor Michele Galietta

Psychologists are commonly called upon to conduct evaluations of a defendant's competency to stand trial. Under *Dusky v. United States* (1960) the legal criteria for competency to stand trial were enumerated and since then, a number of standardized assessment instruments that aim to assess those criteria have been developed, each with its own noted strengths and weaknesses. Although there are several instruments available to aid clinicians in these types of evaluations, only three include screens for feigning, and only one assesses for feigned cognitive impairment. In the current research an instrument was constructed to assess for competence related knowledge, while also incorporating several logical and statistical methods to assess for a feigned lack of knowledge of the legal system, including forced-choice testing, floor effect strategies, and completion time methodologies. The Factual Understanding Instrument (FUI) was constructed over five studies. Studies 1-3 involved instrument construction and included a review of the literature, a critical incidents phase with experts in the field, and item construction. Studies 4-5 focused on item evaluation and included an expert review of the constructed items and the pilot testing of the FUI in a simulation study with unimpaired college students. In study 5, multiple statistical analyses were conducted to evaluate the FUI items and the various feigning

detection strategies. In this sample reliability of the FUI was high. Items were relatively easy for honest responding participants, with many scoring near perfect. Feigning participants did not score as low as would be predicted by symptom validity testing, as responses varied from less than 50% correct to values seen in honest responders. Intelligence level, item difficulty, and response condition were found to be significant predictors in responses to FUI items. Completion time was not supported as a feigning detection method as hypothesized, however, alternative interpretations of the theory are offered. Further research on the FUI with a known-groups sample in forensic settings is needed to establish a floor value, to further evaluate item performance, and to improve the external validity of the current research. Research methodologies and future directions are offered.

ACKNOWLEDGEMENTS

I would like to thank all of my research mentors who have been supportive throughout my graduate school career. It is through your knowledge that I have become the professional I am today. I also want to thank my family and friends who have been solid supporters of all my endeavors and who helped me to believe that I have a wealth to offer this world. Special thanks go to those I came into contact with before I began my career in mental health; your lives have impacted me more than words could ever say.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Construction of a Forced-Choice Task for the Assessment of Factual Understanding and Feigning in Competency to Stand Trial Evaluations

Psychologists are commonly called upon to aid the judicial system in making legal decisions by providing expert opinions. One of the most common psycholegal issues is competency to stand trial (Grisso, 2003; Roesch & Golding, 1980; Rogers, Grandjean, Tillbrook, Vitacco, & Sewell, 2001), which addresses the question of whether a defendant who has been charged with a criminal offense has the capacity to proceed in the criminal adjudication process. More specifically, competency refers to whether a defendant has the present ability and capacity to understand the nature of the legal process as it applies to themselves, and whether the defendant has the ability to function in the legal process by working with his or her attorney (Melton, Petrila, Poythress, & Slobogin, 2007). If there is suspicion that a defendant may have deficits that would interfere with his or her ability to participate adequately in the defense, the law requires that the defendant's competency be evaluated before proceeding as a means of protecting one's right to a fair trial (Grisso, 2003).

Dusky v. United States (1960) was the landmark case in which the legal criteria for competency to stand trial were enumerated. In *Dusky v. United States* (1960), the Supreme Court stated that a simple review of the defendant's mental status by the judge was not sufficient and that "the test must be whether he has sufficient present ability to consult with his lawyer with a reasonable degree of rational understanding and whether he has a rational as well as factual understanding of the proceedings against him" (p.789). In order to be rendered incompetent to stand trial, it must be determined that a defendant has a significant deficit in one or more of the areas (Drob, Berger, & Weinstein, 1987).

While it has been the subject of debate, psychologists have generally come to conceptualize competency to stand trial as comprising three related abilities: having the ability to consult with counsel, having a factual understanding of the court proceedings, and having a rational understanding of the proceedings (Otto, Poythress, Nicholson, Edens, Monahan, Bonnie, et al., 1998; Rogers et al., 2001; Rogers, Jackson, Sewell, Tillbrook, & Martin, 2003). The most basic of the three related prongs is that a defendant must have a fundamental understanding of the legal system, such as knowledge of the various participants in the trial process and their roles in the courtroom, the various ways a defendant can plea, and potential penalties (Grisso, 2003). The ability to consult with counsel, which is also referred to as the reasoning component of competency, encompasses the defendant's ability to understand that his or her attorney is there to provide assistance and to help prepare the best defense against the charges (Hoge, Poythress, Bonnie, Monahan, Eisenberg, & Feucht-Haviar, 1997). Additionally, this component involves ensuring the defendant is able to provide relevant information to the attorney regarding the alleged criminal event, and that he or she can manage the demands of the trial process. The last prong, rational understanding, also referred to as the appreciation factor, refers to the defendant's ability to apply basic legal knowledge to his or her own situation. As such, rational understanding goes beyond the basic understanding of the courtroom and its procedures to understanding that he or she will be going to trial, is facing criminal charges, and that the outcome of the trial will have an impact on this individual's life (Grisso, 2003).

This conceptualization of competency to stand trial has served as a means of operationalizing the *Dusky v. United States* (1960) standard, which ideally allows for consistent evaluation of the abilities most important to assess when determining one's competency status. While these three competence-related abilities should be those considered in all cases, it is

important to note that a defendant's abilities should always be considered in light of his or her specific case, as the anticipated demands of the defendant's trial are directly related to the level of functioning necessary for the defendant to participate adequately during the trial (Grisso, 2003; Zapf & Roesch, 2000).

In cases where the mental state of a defendant is called into question, the defendant is generally referred to a mental health professional for an evaluation to aid the court in determining his or her competence to stand trial. It has been estimated that 60,000 competency to stand evaluations are conducted annually in the United States, therefore, this is an area of utmost importance (Bonnie & Grisso, 2000). As such, the courts rely heavily on psychologists to assess competence-related abilities in defendants and to provide recommendations regarding the competency status of each individual. Research has consistently shown a very high agreement rate between the recommendations provided by mental health professionals and the determination of competency made by the courts (Freckelton, 1996; Hart & Hare, 1992; Zapf, Hubbard, Cooper, Wheeles, & Ronan, 2004). Thus, accuracy in these evaluations is of the utmost importance.

Assessment of Competency to Stand Trial

The concept of competency to stand trial dates back to English common law which held that an individual who was "mad" could not conduct his or her own defense (Blackstone, 1783). As such, historically, severe mental illness has been equated with incompetence and evaluations by mental health professionals typically focused on diagnosing a mental illness that would be likely to interfere with competence (Grisso, 2003). Currently, the presence of a mental illness is but one factor to be considered, and instead evaluations focus on the identification of how specific deficits (i.e., psychiatric symptoms or cognitive deficits) interfere with the three

competence-related abilities. In a well done competency evaluation, the focus should be on how a given deficit or symptom affects the defendant's factual and rational understanding of the proceedings or his or her ability to consult with defense counsel.

Traditionally, clinicians have relied on unstructured interviews and traditional clinical measures to aid in the assessment of a defendant's competence-related abilities. Although an interview can be a direct route to obtaining information, the format and content of an interview may vary between examiners and the information obtained in an interview is dependent on the examiner asking all the necessary questions in order to assess the defendant's abilities. Likewise, traditional clinical measures may be useful for gathering information; however, they are limited as they do not speak to the specific legal issue of competency to stand trial. Therefore, to assist evaluators in conducting valid and reliable assessments, instruments that specifically tap into the three competence-related abilities have been developed. The aim of these competence assessment instruments is to provide mental health professionals with a standardized manner in which to gather information that is legally and clinically relevant to competency recommendations made by practitioners and competency determinations made by the courts (Rogers et al., 2001).

Competence Assessment Instruments

The first attempts at constructing instruments to aid in the assessment of competency to stand trial led to the development of checklists and sentence completion tasks that generally tapped into the factual understanding aspect of competence. Instruments such as Robey's checklist (Robey, 1965), the Competency Screening Test (CST; Lipsit, Lelos, & McGarry, 1971), and the Competency to Stand Trial Assessment Instrument (CAI; McGarry, & Curran, 1973) were of the first developed. Robey's checklist included a basic list of items evaluators

should consider when assessing a defendant's competency, whereas the CST consisted of a sentence-completion task in which the examinee had to fill-in items such as "When I go to court the lawyer will..." While the intention of the CST was to serve as a brief screen of competence-related abilities, the CAI was developed to provide a more comprehensive evaluation. The CAI utilized a semi-structured interview format to assess for competence-related abilities, containing 13 items related to legal issues, such as "appraisal of available legal defenses" and "appreciation of charges." Subsequent to the construction of these instruments, the Georgia Court Competency Test (GCCT; Gothard, Rogers, & Sewell, 1995) was developed. The GCCT included a semi-structured format of assessing legally relevant knowledge, such as knowledge of courtroom figures, and the charges the defendant is facing, but it also added a visual identification task to aid in the assessment of the defendant's basic legal knowledge. Even though these "first generation" instruments provided increased structure for clinicians in their evaluations, criticisms of these instruments were they did not address the rational understanding prong (Cruise & Rogers, 1998; Pinals, Tillbrook, & Mumley, 2006; Zapf & Viljoen, 2003) and they had weak psychometric properties (Pinals et al., 2006; Rogers, Ustad, Sewell, & Reinhardt, 1996; Zapf & Viljoen, 2003).

More recently, empirically validated and standardized instruments that attempt to assess all relevant aspects of *Dusky v. United States* (1960) have been developed. Examples of these "second generation" instruments include the Fitness Interview Test-Revised (FIT-R; Roesch, Zapf, Eaves, & Webster, 1998), the MacArthur Competence Assessment Tool-Criminal Adjudication (MacCAT-CA; Poythress, Nicholson, Otto, Edens, Bonnie, Monahan, et al., 1999), the Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR; Everington & Luckasson, 1992), and, most recently, the Evaluation of Competency to

Stand Trial-Revised (ECST-R; Rogers, Tillbrrok, & Sewell, 2004). The FIT-R utilizes a structured interview format to assess aspects of competence-related knowledge under Canadian law and although constructed for use in Canada, the FIT-R can be adapted for use in the U.S. as it lends itself to an evaluation of the three aspects of *Dusky v. United States* (1960) (Roesch et al., 1998). The MacCAT-CA, a standardized instrument, consists of 22 structured items covering the three prongs of competency. In addition to assessing for the three prongs, the MacCAT-CA includes a unique "teaching" process, to not only assess a defendant's existing legal knowledge, but also his or her capacity to attain such knowledge (Grisso, 2003; Zapf & Viljoen, 2003).

A structured instrument developed for the assessment of competence-related abilities in defendants diagnosed with mental retardation, the CAST-MR, includes the novel approach of utilizing a multiple-choice format to assess for a defendant's legal knowledge (Everington & Luckasson, 1992). Additionally, the CAST-MR includes an open-ended section in which the defendant is questioned on the specifics of his or her own legal situation. The ECST-R is a standardized, semi-structured interview designed to assess competence-related abilities and more specifically, the ECST-R focuses on assessing the influence of psychiatric symptomology on a defendant's abilities (Rogers, Tillbrrok, & Sewell, 2004). An improvement in all the second generation instruments is that they were specifically constructed to correspond to the legal requirements of competence as delineated under *Dusky v. United States* (1960) (with the exception of the FIT-R) and have been empirically validated in clinical samples (Grisso, 2003; Zapf & Viljoen, 2003).

While these instruments are all commercially available, and all represent advancements in the assessment of those abilities thought to underlie competency to stand trial, there is some data that shows mental health practitioners are not consistently using these instruments in their

evaluations. For example, Borum and Grisso (1995) surveyed 57 psychologists who were Board Certified in Forensic Psychology and found that only 36% reported that they 'almost always' use a forensic assessment instrument when conducting a competency to stand trial assessment and 36% reported that they 'never' use them. Additionally, in an archival study of evaluations of competency to stand trial, Skeem, Golding, Cohn and Berge (1998) found only 25% of the reports they reviewed cited the use of a competency to stand trial instrument. These studies are a bit dated, as they were conducted prior to the release of some of the newest instruments (i.e., MacCAT-CA, ECST-R), and a more recent study of 53 diplomats of the American Board of Forensic Psychology (ABFP) revealed those surveyed reported that they incorporate, with at least some frequency, an average of five different competency to stand trial instruments in their assessment practices (Lally, 2003). Thus, while the frequency of use appears to be increasing, the question remains as to the quality of information gleaned from these instruments and the degree to which practitioners rely upon the data derived from them.

Although the second generation competency to stand trial assessment instruments represent an improvement over those that were previously available, there are inherent weaknesses that likely contribute to their limited use in competency to stand evaluations. One issue that arises is in the practicality of the use of these instruments. In a recent study, Pinals and colleagues (2006) found that they were unable to administer the MacCAT-CA to one third of the defendants undergoing competency to stand trial evaluations because of difficulties defendants had in comprehending the test items and the time consuming nature of the instrument. In addition, they found that mild impairments in attention and abstract reasoning led to low scores that may not equate with absolute deficits in competence-related abilities but that were a reflection of difficulties encountered during test taking. This research suggests that impairments

identified by competency instruments, such as the MacCAT-CA which was developed with a strong theoretical foundation and which is the result of extensive empirical research, may not necessarily equate with the deficits in those abilities needed to proceed to trial, but instead represent deficiencies in test-taking abilities. Therefore, these instruments may not aid the clinician in their overall assessment of a defendant's competency, and could in some cases further confuse the issue.

Another drawback of the available competence assessment instruments is that while the authors of the second generation instruments took into account the rational prong, this aspect of competency is difficult to adequately assess in a structured interview format, and even more difficult to adequately assess in a standardized instrument. In a statistical analysis of the three areas assessed by the MacCAT-CA (understanding, appreciation, reasoning), researchers found that the appreciation factor, which focuses on whether the defendant can think about their legal situation and apply relevant knowledge to that situation, had the lowest factor loading, indicating that the items tended to not adequately tap into this aspect of competency (Zapf & Roesch, 2005). The authors suggested that the reason the appreciation variable loaded the lowest of all three is because of the inherent difficulty in measuring an individual's appreciation of his or her particular situation (i.e., the interaction between understanding and awareness of the legal context) in a standardized interview. Additional studies have supported these findings (Jacobs, Ryba, & Zapf, 2008; Rogers et al., 2001).

In addition to the quantitative studies that support the notion that assessment of the "rational understanding" prong is complex, Morris, Haroun, and Naimark (2004) relied on a qualitative analysis to come to similar conclusions. For instance, in their study, forensic examiners described the rational understanding prong as "unsatisfactorily vague" and "confusing

and ambiguous." Morris and colleagues (2004) found that this aspect was difficult to measure reliably and that there was no consensus on how one should assess these qualities in a defendant.

A third drawback recently identified relates to whether these newer instruments all assess the same three competence-related abilities drawn from *Dusky v. United States* (1960), but do so in different ways, and whether these abilities are measured at the same level. Specifically, Ryba (2007) found that agreement rates on the MacCAT-CA and the ECST-R were only 42% for an overall determination of competence, and ranged from a high of 57% agreement for scales that assess factual understanding to a low of 35% for scales assessing appreciation (reasoning or ability to consult with counsel = 53%). While additional studies with these, as well as with other second generation instruments, need to be conducted, these findings indicate that while the tests are designed to tap into the same abilities, they appear to do so in different ways and at different levels.

In sum, it seems that even though there are several options for the use of standardized competence assessment instruments, there is not yet consensus on the utility and practicality of their use, whether they adequately measure the areas that tap into abstract reasoning and the ability to apply information to one's own legal situation, or whether they make similar determinations regarding the presence of deficits in competence-related abilities. Therefore, practitioners must proceed with caution when using these instruments.

Feigning and Competency to Stand Trial

In addition to the inherent weaknesses of these competence assessment instruments, a common critique is that, with only a few exceptions, these measures are susceptible to feigning (Everington, Notario-Smull, & Horton, 2007; Gothard, Rogers, & Sewell, 1995; Gothard, Viglione, Meloy, & Sherman, 1995; Rogers, Sewell, Grandjean, & Vitacco, 2002; Vitacco,

Rogers, Gabel, & Munizza, 2007). One of the basic assumptions underlying the interpretation of any psychological test is that the person is performing to the best of his or her ability (Rogers, 1997; Tombaugh, 2002). However, in a forensic setting the clinician must be alert to the possibility that a defendant may attempt to present in a manner that is inconsistent with his or her actual abilities. When necessary, professionals may utilize separate instruments to assess for feigning; however this adds additional time to the evaluation and the relationship between these instruments and competence assessment is not necessarily clear.

Only two of the competence assessment instruments described above have incorporated measures of feigning into the test: the GCCT and the ECST-R. The GCCT includes a brief eight-item screen for psychopathology. These items were purposely constructed to have either a relatively non-bizarre quality (e.g. association of judge's black robes with black magic) or bizarre qualities (e.g. confusion of criminal charges with electrical impulses and credit card debts). The items were designed to distinguish between potential feigners and those with bona fide mental illness via patterns of responses on the different types of items (Gothard et al., 1995). Similar to the GCCT, the ECST-R (Rogers, Jackson, Sewell, & Harrison, 2004) includes a brief screen for feigned mental illness. The items on the ECST-R assess for psychotic and non-psychotic symptoms related to problems a defendant may have within the context of a trial and elevations on these scales indicate the presence of overreporting of symptoms of mental illness. The authors of the GCCT and ECST-R recommend cut-scores for their feigning scales and when a defendant endorses a number of items above this cut score, further assessment for feigning is suggested.

The feigning scales on the ECST-R have shown promising results as a useful screen for feigned psychopathology (Vitacco, Rogers, & Gabel, 2009; Vitacco et al., 2007). However, the

scales built into both the GCCT and ECST-R are focused on the assessment of feigned mental illness. Persons undergoing competency to stand trial evaluations may attempt to feign other impairments, such as feigned deficits in memory, attention, knowledge or other cognitive-based approaches. Recent research has highlighted the importance of assessing for these feigning presentations (Vitacco et al., 2009).

Malingering and Feigning

Malingering refers to the deliberate fabrication or gross exaggeration of psychological or physical symptoms for the fulfillment of an external goal (American Psychiatric Association, 2000). The *Diagnostic and Statistical Manual of Mental Disorders-4th Ed. Text Revision* (DSM-IV-TR; APA, 2000) states that "malingering should be strongly suspected if any combination of the following is noted: 1) medicolegal context of presentation (e.g., the person is referred by an attorney to the clinician for examination), 2) marked discrepancy between the person's claimed stress or disability and the objective findings, 3) lack of cooperation during the diagnostic evaluation and in complying with the prescribed treatment regimen, and 4) the presence of antisocial personality disorder" (p. 739). In a forensic setting there is typically a presumption of an intended secondary gain in defendants, such as avoiding trial or punishment. But, because a diagnosis of malingering specifically requires the identification of an anticipated secondary gain, the term feigning is used to refer to the deliberate exaggeration or creation of symptoms when there is no clear reference to an external goal (Rogers & Bender, 2002). Therefore, psychological instruments developed to detect 'malingering' actually attempt to detect 'feigning,' as no instrument can also identify the secondary gain that may result from an altered presentation. Because of this reasoning, the term feigning shall be used in the remainder of this paper rather than malingering.

Feigning can encompass different domains of portrayed symptomotology, such as feigned psychiatric disorders, cognitive impairment, and medical illness (Rogers & Bender, 2002; Rogers, Salekin, Sewell, Goldstein, & Leonard, 1998). Psychologists are typically confronted with feigned psychiatric disorders or cognitive impairment (Graue, Berry, Clark, Sollman, Cardi, Hopkins, et al., 2007) and there is evidence that within the forensic context feigned psychiatric disorders tend to be the most salient (Rogers et al., 1998). Generally, those attempting to feign tend to over exaggerate deficiencies, present with uncommon symptoms, and present in a more extreme manner than would a person who truly suffers from a psychological or neurological disorder (Gothard et al., 1995; Kucharski, Ryan, Vogt, & Goodloe, 1998).

While the exact prevalence of malingering will never be truly known (Rogers, 1997), experts suggest that it occurs in 13-17% of forensic cases (Heinze & Purisch, 2001; Rogers et al., 1998; Rogers & Bender, 2002), with up to a 21% rate of probable malingering in forensic hospital settings (Vitacco et al., 2007). When a defendant is suspected of feigning during a competence evaluation and a secondary gain can be identified, an evaluator may conclude the defendant is malingering. Defendants undergoing competency to stand trial evaluations are at a high risk for feigning because they may attempt to feign incompetence to delay or avoid a trial, to avoid legal punishment, or to be transferred to a hospital rather than a jail or prison (Resnick, 1997). It has been suggested that clinicians conducting competency evaluations should adopt a relatively low threshold for suspecting feigning, but at the same time should also not be overly hasty in reaching the decision due to the negative effects that can result when someone who is genuinely disordered is labeled as feigning (Rogers, 1997).

In the assessment of feigning, sole reliance on clinical judgment is inadvisable due to the unreliability of the judgments (Hickling, Blanchard, Mundy, & Galovski, 2002; Melton et al.,

2007; Rogers & Bender, 2002). Although the clinical identification of feigning is important (Kucharski et al., 1998), it is desirable to have support from multiple sources when making a determination and the inclusion of a standardized assessment instrument can often aid the clinician and provide verification for suspicions raised during a clinical interview (DeClue, 2002).

Instruments Used to Assess Feigning

There are many ways standardized assessment instruments can help practitioners to identify feigning. Traditional clinical personality assessment instruments are often helpful in detecting feigning through the assessment of a defendant's response style. For example, the Minnesota Multiphasic Personality Inventory-Second Edition (MMPI-2; Butchner, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and the Personality Assessment Inventory (PAI; Morey, 1991) include validity indices that can offer information as to whether the person is responding consistently and within a range expected for different groups of individuals who take that test. Additionally, these instruments have scales that can identify when a respondent is intentionally overreporting symptoms in an attempt to appear more pathological than they actually are. Although these personality measures can provide useful information, they are quite time consuming, may not be practical to administer during a competency evaluation, and were not specifically developed for use in competency evaluations. Therefore, less time consuming and more practical means of assessing feigning are necessary. Currently, there are several empirically validated and standardized instruments that have been specifically designed to detect feigning and these tests can generally be categorized as either assessing for feigned psychiatric impairments or feigned cognitive impairments.

Assessment of feigned psychiatric impairment. A number of instruments have been specifically designed to assess for feigned psychiatric disorders. One of the most widely used is the Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992). The SIRS is a 172-item structured interview that assesses for symptoms that are infrequently endorsed in genuine clinical populations, atypical symptoms that are very likely to be untrue, and an overall endorsement of symptoms that is higher than those found in genuine psychiatric patients. Research supports the reliability and validity of the use of the SIRS in the detection of feigned psychiatric disorders (Rogers, Gillis, Dickens, & Bagby, 1991; Rogers, Jackson, Sewell, & Salekin, 2005). And although the SIRS was primarily designed to assess for feigned psychiatric disorders, the instrument includes a brief measure of feigned cognitive impairment, the IF scale (Improbable Failure), which consists of asking the examinee to identify a grammatical opposite and a homonym for a short list of words.

The Miller Forensic Assessment of Symptoms Test (MFAST; Miller, 2001) is another instrument used to assess for feigned psychiatric disorders. The MFAST is a short screening instrument that consists of 25 items similar to those found on the SIRS but can be used in cases where a quick screen for malingering is desired. Research supports the utility of the MFAST in forensic settings for the detection of feigned psychiatric symptoms (Jackson, Rogers, & Sewell, 2005), although it should be noted that short screening measures such as the MFAST should be used with caution as a positive finding on this type of test should not be equated with feigning. Most researchers suggest a two stage process in which a brief screen is used to identify cases of possible feigning by screening in those who are in need of a more comprehensive assessment and screening out those who do not raise red flags in the evaluation (Rogers et al., 2004; Vitacco et al., 2007). For those flagged as possible feigners, a more thorough examination is then required

as the determination cannot be made on the basis of elevated scores on a screening instrument alone. In sum, with these instruments at the disposal of evaluators, one has a number of options should suspicions of feigned mental illness arise during a competency evaluation.

Assessment of feigned cognitive disorders. Because instruments like the MFAST and SIRS are designed specifically to detect feigned psychiatric symptoms, they are not likely to be very sensitive to other types of feigning, such as feigned cognitive impairment (Alwes, Clark, Berry, & Granacher, 2008). Additional measures have been developed that focus on the assessment of feigned cognitive deficits. Examples of such instruments include the Portland Digit Recognition Test (PDRT; Binder & Willis, 1991), the Validity Indicator Profile (VIP; Frederick, 1997) and the Test of Memory Malingering (TOMM; Tombaugh, 1996). The PDRT and TOMM are instruments used in the assessment of feigned memory impairment and are similar in that a series of visual stimuli are presented to the examinee and a recognition task of the presented items follows. While the PDRT utilizes the presentation and recall of numbers, the TOMM consists of line drawings of commonly known objects. The VIP differs from these instruments in that it is used to assess verbal and nonverbal abilities in an examinee. This instrument utilizes sophisticated statistical techniques to assess the examinee's motivation and level of effort regarding their response to the test items. The Structured Inventory of Malingered Symptomology (SIMS; Smith & Burger, 1997), another feigning detection instrument, is unique in that it assesses for both feigned cognitive impairment and feigned mental illness, and recent research support its utility in forensic settings (Jelicic, Hessels, & Merckelback, 2006).

Though the instruments designed to detect feigned cognitive impairment have proven useful in forensic settings, and examiners may use these tests in their competency evaluations when cognitive feigning is suspected, there has been only one attempt to construct a competency

assessment instrument which incorporates the detection of cognitive feigning directly into the test. The Test of Malingered Incompetence (TOMI; Colwell, Colwell, Perry, Wasieleski, & Billings, 2008) was the first instrument to incorporate a method of assessing for feigned cognitive impairment. Although this instrument was a useful first step, the TOMI is not commercially available and only one peer-reviewed article has been published on its psychometric properties. Furthermore, the TOMI includes only one method for assessing for feigning (i.e., symptom validity testing). It stands to reason that an instrument could be constructed that incorporates multiple feigning detection techniques into the assessment of competence related knowledge. Incorporating these techniques could potentially aid examiners by increasing both the accuracy (i.e., detecting feigning through multiple means) and efficiency (i.e., save time by using only one instrument) of their evaluations.

Techniques Used to Assess Feigning

There are many techniques that can be used when constructing an instrument for the assessment of feigning. These techniques include the use of various statistical procedures and logical analyses and each test designer may rely on one or many of these methods in the construction of their instrument. Examples of these are symptom validity testing, floor effect strategy, performance curve classification, assessment of completion time, and atypical symptom presentations (Rees, Tombaugh, Gansler, & Moczynski, 1998; Rogers & Bender, 2002; Rogers et al., 2002). Instruments used to assess feigned psychiatric disorders typically assess for improbable combinations of endorsed symptoms (atypical symptom presentation), whereas measures of feigned cognitive impairment typically employ symptom validity testing, floor effect strategy, performance curve classification, and completion time methodologies. Though each technique has its inherent strengths and weaknesses, the focus herein will be on symptom

validity testing, floor effect strategy, and completion time methodologies in the assessment of feigning.

Symptom Validity Testing (SVT). SVT has been used to assess feigning in instruments such as the TOMM (Tombaugh, 1996) and the PDRT (Binder & Willis, 1991) and is often based on the forced-choice recognition paradigm. Pankratz, Fausti, and Peed (1975) initially proposed the use of a two-alternative, forced-choice task to assess for deception in a case study of a patient with suspected malingered sensory deficits. Pankratz and colleagues (1975) reported that the patient was suspected of feigning deafness in order to gain compensation for his disabilities and when they presented him with a task in which 100 trials of audio tones were presented, the patient correctly identified only 36 of the tones. The investigators rationed that by chance alone a person should correctly identify 50 of the 100 trials, and that the probability of the patient obtaining such a low score was .004. Thus they were able to conclude that he was actively choosing the wrong responses.

Of the forced-choice recognition tasks, SVT is the most commonly used approach (Weinborn, Orr, Woods, Conover, & Feix, 2003). This type of strategy for detecting feigned impairment refers to an improbable failure rate based on statistical probability in that there is a known chance level of performance, (i.e. 50%, as was with the original Pankratz and colleagues' procedure). A common feature of the SVT method, as with a test like the TOMM, is that the test appears more difficult than it actually is. This increases the likelihood that dishonest responders will perform at a below chance level (Bianchini, Mathias, & Greve, 2001; Tombaugh, 1997).

The rationale behind SVT is that a person must be capable of scoring significantly above chance in order to score significantly below chance (Hiscock & Hiscock, 1988). Hiscock, Branham, and Hiscock, (1994) argue that severely impaired individuals should be able to score at

or above chance level (50%) and that below-chance performance on a forced choice test "is so compelling an indicator of dishonest responding that it can serve as a criterion for selecting malingerers in clinical studies" (p. 96).

SVT has been implemented widely in neuropsychology in the assessment of feigned cognitive impairment (Tombaugh, 1997; Weinborn et al., 2003; Willison & Tombaugh, 2006) and promising results have been found with SVT in distinguishing simulated feigners from controls (Tan, Slick, Strauss, & Hultsch, 2002), coached from uncoached feigners in the simulation of head injury (Tombaugh, 1997; Weinborn et al., 2003; Willison & Tombaugh, 2006), and those who feigned lack of remote memory for an alleged crime (Denney, 1996). Based on the positive findings in using this procedure, the forced-choice task has been highly recommended for the assessment of response bias (Millis & Volinsky, 2001).

There are, however, limitations of SVT; namely that such tests may be transparent to the malingerer who may perform poorly but not to the level as to score below chance (Hiscock et al., 1994). Weinborn and colleagues (2003) note that when this occurs, it will result in unacceptably low levels of sensitivity and they suggest that a more accurate alternative is to use cutoff scores derived from the performance of genuinely impaired persons. Thus, many tests have also incorporated a floor effect strategy as well as detecting below chance performance.

Floor Effect Strategy (FES). The floor effect strategy refers to the comparison of scores obtained by those suspected of feigning or altering their performance on the test, to the scores of those who are known to be demonstrably impaired (i.e. the "floor"). Once the floor is established in a sample of persons who are genuinely impaired, a non-impaired individual's performance that falls below this floor can be deemed suspicious (Frederick, 2000). Tombaugh (1997) utilized a floor effect strategy in the construction of cutoff scores for the TOMM. For example, in a

group of participants with cognitive impairment, traumatic brain injury, or dementia, it was rare for their performance on the TOMM to fall below 45 out of 50 correct. Thus, on the basis of establishing a floor performance of those with true cognitive impairment, the score of 45 out of 50 produced a high level of specificity with cognitively impaired and demented persons with correct classification of not feigning 91% of the time. In addition, in a follow-up simulation design, none of the feigners received a score higher than 42 out of 50, suggesting that the cutoff score yielded a sensitivity and specificity rate of 100%. Similar results were obtained in follow-up validation trials of the TOMM (Rees et al., 1998). This technique has proven useful in the detection of feigning, and works well in combination with SVT based on a forced-choice paradigm.

Completion Time (CT). Another method that can aid in the assessment of feigning is completion time (CT). In addition to measuring the correctness of a given response, the amount of time that is required for a feigner to plan and execute a response can also give an indication of feigning. That is, a person who is trying to alter his or her performance must think about how to respond and, as a result, the time it takes that person to complete the test becomes slower than a person who is responding automatically and genuinely (Reicker, 2008; Willison & Tombaugh, 2006). Increased completion time also results from feigners having to focus their attention on how many items they should or should not answer correctly which unintentionally increases the time it takes to complete the test. In an attempt to investigate completion time performance, Rees and colleagues (1998) found that those who were instructed to simulate a cognitive deficit on the TOMM consistently produced longer times. Similarly, Willison and Tombaugh (2006) found that feigners produced longer completion times than did control participants, and longer completion times than those with traumatic brain injury (TBI). More recently, Reicker (2008)

found completion time to be the best predictor of feigned performance on a computerized information processing test (as compared to task complexity predictors).

Although the previous findings apply to the simulation of TBI, the concept of assessing completion time has been applied to other domains of impairment. For instance, this technique has also been found useful on the SIRS; in that feigners took longer to complete the instrument than honest responders (Rogers et al., 1992). Everington and colleagues (2007) measured the length of time it took participants to complete the CAST-MR, comparing those who were responding honestly to those instructed to feign impairment of knowledge. Differences were found between the honest and feigning participants, and although it did not reach statistical significance, feigners not diagnosed with mental retardation took an average of two minutes longer to complete the CAST-MR than did the honest responders. With these findings, it is of interest to continue to investigate completion time as a means of assessing feigned performance in competency to stand trial evaluations with different test formats, as the Everington and colleagues (2007) study was the first study to utilize this methodology. Overall, this is not an area which has been greatly explored in regards to the assessment of competency to stand trial and feigning.

Assessment of Competency to Stand Trial and Feigned Incompetence

As previously discussed, there are several instruments available for the assessment of competence related knowledge and there are a number of additional instruments available for the assessment of feigned incompetence. However, there have only been three attempts to incorporate the assessment of both competence related knowledge and feigning into the same instrument: the GCCT (Gothard et al., 1995), the ECST-R (Rogers et al., 2004), and the TOMI (Colwell et al., 2008). As previously noted, the GCCT includes Atypical Presentation Scales

(ATP) that consist of closed-ended questions specifically related to the courtroom process and that focus on symptoms of psychotic disorders. The concept underlying the ATP scales is that those who are feigning are more likely to endorse bizarre items whereas those who are honest will not. While the idea of assessing for atypical symptom presentation was novel at the time the GCCT was developed, research has shown that these scales have only a moderate sensitivity rate, suggesting that many potential feigners will not be identified (Rogers et al., 2002). Similarly, the ECST-R includes five ATP scales that serve to screen for feigned psychotic disorders in addition to evaluating a person's competence-related knowledge. The ATP scales on the ECST-R assess for the endorsement of psychotic and nonpsychotic symptoms of mental illness, for the endorsement of symptoms that are uncommon in psychiatric populations, and for the endorsement of absurd symptom combinations. These scales are relatively quick to administer and the test authors have found that the ECST-R ATP scales are efficient at identifying cases in which a further comprehensive evaluation of feigning is necessary (Rogers et al., 2002).

The premise of the feigning scales on the GCCT and ECST-R is to detect feigned psychiatric illness; however as previously noted, defendants undergoing competency evaluations may also attempt to feign cognitive deficits in order to appear impaired. In a recent study, Vitacco and colleagues (2009) administered the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999), SIRS, and ECST-R to a sample of inpatient pretrial defendants and found that those in the probable feigning group produced significantly lower estimated IQ scores on the WASI than did those in the genuinely disordered group. Furthermore, participants in the probable feigning group scored significantly lower on the IF (Improbable Failure) scale of the SIRS than did those in the genuinely disordered group. These results provide further support for

the notion that individuals may attempt to portray cognitive deficits while undergoing a competency to stand trial evaluation.

Since not all feigning attempts consist of feigned psychopathology, screens for feigned cognitive deficits such as memory impairment or feigned deficits in attention or concentration may be useful in some competency to stand trial evaluations. However, a screen for attempts to feign a general lack of knowledge of the legal system would be useful in all competency to stand trial evaluations. Such a test could tap directly into the level of effort and honesty in an individual's presentation and could also speak to one of the three relevant competence-related abilities that must be assessed. Thus, a test that measures one's legal knowledge (i.e., level of factual understanding) and also identifies attempts to purposefully appear ignorant of knowledge related to the legal system could prove very useful for clinicians who conduct competency evaluations.

In a recent study, Everington and colleagues (2007) took a pre-established measure of competency to stand trial, the CAST-MR, and assessed whether defendants diagnosed with mental retardation were successful in faking impairment. It was found that defendants with mental retardation, when told to pretend that they did not know the answers to the test, scored similar to defendants without mental retardation who were also told to fake a disability. This study is of interest because it represented the first attempt to assess for a feigned lack of legal knowledge. While their results were only moderately promising, had they incorporated additional feigning detection strategies into their evaluation, this method of feigning detection could potentially prove highly useful.

As previously noted, the rational understanding prong of competency to stand trial is difficult to measure and attempts to construct a standardized instrument that captures this

cognitive process have not been fully successful (Jacobs et al., 2008; Rogers et al., 2001; Zapf & Roesch, 2005). The area researchers have reliably been able to assess, and which serves as the most basic component of competency, is a factual understanding of basic legal knowledge. Even though some defendants may have a genuine lack of knowledge, there may be some who will intentionally feign ignorance. An instrument that can aid in identifying individuals who intentionally feign while simultaneously assessing for legal related knowledge could prove to be very useful. Such an instrument can also be practically useful for clinicians if it is quick to administer and easy to interpret.

As previously mentioned, the TOMI (Colwell et al., 2008) is the first instrument to assess for legal related knowledge that also includes a screen for feigned cognitive impairment. The TOMI is a 50-item test that utilizes a forced-choice testing paradigm for the assessment of general knowledge (TOMI-G) and for legal knowledge (TOMI-L). The authors of the TOMI pilot tested the instrument in a student sample and in a small sample of forensic patients. In the forensic sample, the TOMI-L performed better than the TOMI-G, with feigning participants scoring significantly lower than the honest participants on the measure of legal knowledge. While these preliminary results give support for the use of the TOMI-L, this instrument has several limitations. First, the instrument has an 8^{th}-9^{th} grade reading level, which may exceed the capacities of forensic populations who generally have lower reading levels and are less educated than non-incarcerated persons (Bates, Davis, Guin, & Long, 1992; Heckel, Allen, Andrews, Roeder, Ryba & Zook, 1989; Kerridge, 2008). Secondly, the TOMI-L incorporates only one type of feigning detection strategy and does not incorporate other measures such as completion time. Third, the content of the TOMI-L includes items that are not directly reflective of the construct of factual knowledge of the legal system. That is, several items appear marginally related to the

legal knowledge necessary for a defendant to proceed to trial (e.g., knowledge of what a gag order is, what a frisk is, what badgering a witness means). Lastly, as previously noted the TOMI is not commercially available for use in evaluations and there has only been one peer-reviewed research article published on its use.

Statement of the Problem

To date, no attempts have been made to directly incorporate multiple strategies to detect a feigned lack of knowledge into an instrument that is designed to assess competence related abilities. It is proposed that an instrument designed to assess legal knowledge, based on a strong definition of the construct of competence related knowledge, could also screen for feigning by using multiple detection strategies (i.e., symptom validity testing, floor effect, and completion time), and that such an instrument could prove highly useful and practical in the assessment of the factual understanding prong of *Dusky v. United States* (1960). A test of this nature could assist evaluators in screening out those who may feign incompetence due to impaired legal understanding. Thus, an instrument aimed at assessing factual understanding that can also serve as a screen for feigned lack of knowledge of the legal system, that it is quick to administer and relatively easy to interpret, would be valuable in assisting clinicians in the assessment process.

The purpose of the current research was to conduct the preliminary steps toward constructing such an instrument. It is important to note that the purpose of the instrument was not to serve as a full competency to stand trial assessment instrument, as this instrument would be insufficient to capture all three prongs, but rather to provide information regarding the defendant's factual knowledge of the issues pertinent to competency to stand trial and also to have the face validity of a competency measure. Although it is pertinent that the instrument accurately assess factual understanding, a primary purpose of the proposed instrument is to serve

as a screen for feigning. Thus, the instrument is intended to serve as a screen for competency to stand trial knowledge and for feigned incompetence due to impaired understanding and knowledge of the legal system.

The development of the instrument followed the guidelines set forth in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) in that careful consideration was given to test construction, including issues of test validity and reliability. The construction of the instrument, herein referred to as the Factual Understanding Instrument (FUI), occurred over five studies. Studies 1-3 involved instrument construction. Study 1 included a review, coding, and analysis of the available literature on competency to stand trial. Study 2 included a critical incidents phase in which experts in the field were interviewed about the information they typically assess defendants for during a competency evaluation. Study 3 focused on the delineation of item specifications and subsequent item construction. Studies 4-5 addressed instrument evaluation. In Study 4 the constructed FUI items were reviewed by experts and the instrument was further refined. In Study 5 the FUI was pilot tested in a simulation study with unimpaired college students.

Study 1

The first step in conceptualizing and creating the FUI was to define the construct to be measured. The construct of factual knowledge of the legal system was defined by means of an extensive review and coding of the existing competency to stand trial instruments and the literature on competency to stand trial. The construct was also defined by incorporation of critical incidents (Study 2).

Method

A Psych Info search was conducted in which the words "competence/competency to stand trial" were entered. All articles were reviewed that addressed the topic. However, only a fraction of those reviewed included a clear definition and conceptualization of the construct of factual knowledge of the legal system. Testing manuals and currently available competence instruments were also reviewed for a definition of factual understanding of the legal system and for items currently used to tap into the construct.

As materials were reviewed, data was coded. The coding system consisted of definitions of the units of material to be analyzed (i.e., the words other researchers have used to conceptualize factual understanding of the legal system [e.g., defense lawyer, judge, jury, etc.]), exhaustive categories of classification (i.e., coding units defined by the primary investigator), and rules for applying the coding system (i.e., the degree to which a given statement fit into the defined unit). Coding units were grouped together with other units which shared common attributes to create categories and all categories were mutually exclusive. Upon completion of this categorization task, each category was assigned a code number. Subsequently, each time a category appeared in a reviewed article, test manual, or instrument the occurrence was recorded

in the given cell of a Microsoft Excel spreadsheet. This process was continued for each data source. After all data were coded and recorded, frequencies were calculated.

Results and Discussion

In total, 25 articles, instruments, and testing manuals were reviewed and coded. A total of 58 categories were defined through his process. The categories with frequencies greater than or equal to five are presented in Table 1. The category with the highest frequency was the defendant's knowledge of the nature/severity of their current charges, followed by knowledge of the various courtroom figures.

Table 1

Frequencies of Factual Understanding Variables Defined in the Literature (With Frequency Greater Than or Equal to 5)

Variable	Frequency
Nature/severity of current charges	24
Role of defense attorney	20
Role of jury	20
Role of prosecuting attorney	19
Range and nature of possible penalties	18
Role of judge	17
General court proceedings/process of trial	16
Various legal pleas (guilty/not guilty)	15
Role of witness	14
Role of defendant	11
Consequence of guilty plea/conviction	11
Adversary nature of court process	10
Defendant understands he/she is accused of a crime	7
Plea bargain	6
Waive rights	6
Testify	5

From these results hypotheses were made about the information that would be covered by FUI items (e.g., knowledge of the types of charges, courtroom figures, types of legal pleas). The results also led to hypotheses about the data that would be gathered from Study 2. It was predicted that the results from Study 1 would be similar to the categories defined in Study 2, but that greater variability and depth would be gained in Study 2.

Study 2

In addition to a review of the available literature and existing competency to stand trial instruments, the second step in defining the construct of factual knowledge of the legal system was to incorporate critical incidents.

Method

"Critical incidents" refers to a procedure for obtaining facts about a given situation, classifying those facts, and making inferences from the information. This study utilized expert interviews as the format for obtaining critical information on the construct of competency to stand trial.

Participants. Ten mental health professionals with knowledge and experience in the field of conducting competency to stand trial evaluations were contacted to participate in a brief and informal interview. Participants in Study 2 comprised a sample of convenience and included individuals known to the primary investigator and/or the research advisors. Participants were contacted by phone, in-person, or e-mail and asked to participate in a brief telephone, in-person, or computerized instant message interview.

Procedure and Apparatus. Upon agreement to participate in the interview, participants were provided with a consent form (either in-person, through fax, or e-mail). A mutually convenient time was arranged for the interview. Interviews were focused on the mental health professionals' practices when evaluating a defendant's competence. The first question was, "When conducting a competency to stand trial evaluation, what questions do you regularly ask to gauge the defendant's factual understanding of the legal system?" After responses to this question were collected, the primary investigator asked, "Of those questions you mentioned that you consider pertinent to gauging a defendant's competence, which three do you think are the

most relevant and pertinent? Of those three, can you please rank them in order of importance from one to three?" Interviews lasted approximately 15 minutes. All in-person and telephone interviews were recorded using an Olympus WS-100 Digital Voice Recorder. Audio conversations were transcribed into Microsoft Word by a trained Masters level research assistant.

Interview transcripts were reviewed by the primary investigator, coded, and categorized in a content analysis. Similar to Study 1, the coding system consisted of definitions of the units of material to be analyzed (i.e., the content each mental health professional endorsed to gauge the defendant's factual understanding [e.g., asking the defendant what he or she is charged with]), exhaustive categories of classification (i.e., coding units defined by the primary investigator), and rules for applying the system (i.e., the degree to which a given statement fit into the defined unit). For example, for question #1, the first participant reported that he or she regularly asked the defendant his or her knowledge of the defense attorney, prosecutor, judge, and jury. From this statement four coding units were created for each courtroom representative. This process continued for participant #1. For each subsequent interview, frequencies were recorded for each previously defined coding unit and additional units were created as needed. Similar units were classified together into a smaller number of categories. Next, those items which were reported as most pertinent in an evaluation were classified under one of the categories.

Results and Discussion

In the content analysis for question #1, 121 units were identified. These units were combined into 24 categories and the categories with frequencies greater than or equal to three are presented in Table 2. Knowledge of the various courtroom figures was the highest rated category, followed by knowledge of the legal pleas.

Table 2

Frequencies of Factual Understanding Variables Obtained Through Expert Interviews (With Frequency Greater Than or Equal to 3)

Variable	Frequency
Knowledge of the judge	8
Knowledge of the jury	8
Knowledge of the defense lawyer	8
Knowledge of the guilty/not guilty/no contest pleas	7
Knowledge of charges	6
Knowledge of the prosecutor	6
Knowledge of the adversary nature of the courtroom	6
Knowledge of misdemeanor/felony	5
Knowledge of plea bargains	5
Knowledge of the witness	3
Knowledge of possible sentences	3
Knowledge of the role of other participants in the courtroom	3
Knowledge of courtroom terminology	3

Of the 24 defined categories, the 12 that were rated as most pertinent and critical in gauging a defendant's factual understanding of the legal system are reported in Table 3. The category rated as most pertinent by the greatest number of participants was the defendant's knowledge of their charges.

Table 3

Coded Frequencies of Critical Incidents Through Expert Interviews

Variable	Frequency
Knowledge of charges	9
Knowledge of defense attorney	3
Knowledge of the role of other participants in the courtroom	3
Knowledge of the guilty/not guilty/no contest pleas	3
Knowledge of the adversary nature of the courtroom	3
Knowledge of the general court proceedings	3
Knowledge of possible sentences	2
Knowledge of plea bargains	1
Knowledge of the crime he/she is accused of	1
Knowledge of legal rights	1
Knowledge of the judge	1
Knowledge of the prosecutor	1

All of the categories defined from this phase were combined with the information collected from Study 1 for incorporation into the construct of factual knowledge of the legal system. Many of the categories from this study overlapped with the information obtained from Study 1, suggesting that the construct of factual knowledge of the legal system was relatively stable as it was defined similarly by the mental health professionals sampled in this study and in the existing literature and competence assessment instruments. From the results of Study 1 and 2, the domains to be covered by the FUI items were ascertained. In Study 3 individual FUI items were constructed.

Study 3

After identifying and defining the construct, the third step in the construction the FUI was to generate items based upon several item specifications. Study 3 included two parts, the generation of items based on those specifications and the development of hypotheses related to item attributes.

Method

Item specifications served as a means of selecting the range of items to cover, the format of the items, the reading level, and characteristics of the correct and incorrect responses (Crocker & Algina, 1986; Downing, 2006). These specifications were created and followed during item construction; they are summarized below.

Stimulus attributes. Simple and concise language was applied to all items. The reading level of the items was assessed with the Flesch-Kincaid grade level readability score (Flesch, 1948). Items were written at a 4th grade reading level (Flesch-Kincaid grade level = 4.1, Flesch reading ease = 80.5%). A 4th grade reading level was chosen because this level has been utilized in other standardized measures of personality (e.g., PAI; Morey, 1991). While some personality measures have utilized a higher reading level (e.g., MMPI-2; Butchner et al., 1989), the intended application of the FUI was with forensic populations, who tend to have less formal education than non-incarcerated persons (Bates et al., 1992; Heckel et al., 1989; Kerridge, 2008). During item construction the objective was to simplify difficult terminology and concepts into concise and straightforward language to reduce construct-irrelevant variance that may result from reading difficulties.

The second stimulus attribute was that all items were written in a question format (e.g. “What is a defendant?”). The third specification was that the items have high face validity as a

measure of competence to stand trial such that the questions reflect the construct of factual knowledge of the legal system. The fourth specification was that the items were written to have high content validity in that the construct was well represented in the items. That is, the categories of factual knowledge of the legal system that were identified via the review of literature, competency instruments, testing manuals, and expert interviews in the previous studies were represented in the items.

Response attributes. All response options met several attributes. The first was that each item contained only two answer choices: a correct answer and an incorrect answer. Because the FUI was designed to assess factual knowledge of the legal system as well as feigning, a forced-choice method in which only two response options are presented was used.

Second, incorrect answer choices fell under one of two categories and were either an obviously incorrect answer or a more plausible, subtle incorrect answer. An obviously incorrect answer referred to an answer that a layperson with even the most basic knowledge of the legal system and a most rudimentary level of intelligence would consider obviously wrong. The subtle incorrect answer was one that a layperson with basic knowledge of the legal system would consider plausible and that a person of average intelligence may get incorrect if lacking in legal knowledge. The two incorrect item types will be utilized as a means of establishing a floor effect (to be used in later studies of genuinely impaired individuals) and of differentiating between responders, which is elaborated on below.

Additional response attributes were executed. Specifically, the within-item response choices (correct and incorrect) were of approximately equal length with a comparable number of words in each answer. Additionally, response choices were matched for level of vocabulary difficulty both within and between items. Within-item responses were matched for grammatical

person (first, second, or third), tense (past, present, or future), voice (active or passive), and quantifiers (“all,” “some,” “most,” etc.).

Item Construction. One or two items were constructed for each coding unit collected from the content analysis and critical incidents phase. For each item, a stem and a correct answer choice was created as was either a subtle or an obvious incorrect answer. Additional items were created, as appropriate, in order to have a mix of items with subtle and obvious responses.

Scoring procedure. For the scoring of items, correct answers received one point and incorrect answers received zero points. Items were tallied to produce three scores: a subscale score for the subtle items the participant answered correctly, a subscale score for the obvious items that were answered correctly, and a total score for the number of items correct on the entire instrument (i.e., the sum of the subscales). A time to completion score was also documented, which is the amount of time (in minutes and seconds) that it took the participant to answer all of the items.

Relationship between response style, intelligence level and item type. One of the intentions of constructing the answer choices in the manner delineated above, with some items having the correct answer and an obvious incorrect answer, and some items having the correct answer and a subtle incorrect answer, was to test the hypothesis that a relationship between intelligence level, response style (honest vs. feigning), and item type (subtle vs. obvious) exists. Table 4 illustrates the predicted relationship between the three variables.

Table 4

Hypothesized Relationship Between Intelligence Level, Response Style, and Item Type

	Intelligence Level		
		Average/Above Average	Below Average/Borderline
Response Style	Honest	O-most correct S-most correct	O-most correct S-some incorrect
	Feign	O-most correct S-most incorrect	O-most incorrect S-most incorrect

Note. O = Obvious Factual Understanding Instrument (FUI) items, S = Subtle Factual Understanding Instrument (FUI) items. Average/Above average intelligence refers to levels which fall no more than one standard deviation below the mean or that fall above the mean. Below average/Borderline intelligence refers to scores that are more than one standard deviation below the mean.

It was predicted that participants with average to above average intelligence (i.e., intelligence scores which fall no more than one standard deviation below the mean or with scores at or above the mean) who responded honestly to the test items, would correctly answer all of the obvious items as well as the majority of the subtle items. Participants who respond honestly and that have low average to borderline levels of intelligence (i.e., intelligence scores which fall more than one standard deviation below the mean) were predicted to correctly answer most of the obvious items but to correctly answer fewer subtle items. Additionally, it was predicted that participants who have average to above average intelligence, but who dissimulated their response style (feigners), would correctly answer most of the obvious questions but would incorrectly answer most of the subtle items as a means of attempting to appear impaired but not so impaired as to not be believed; while participants with low average to borderline levels of intelligence who

feign were predicted to incorrectly answer most of the obvious and subtle items as a strategy of attempting to incorrectly answer the majority of the items.

Results and Discussion

From the coding units a large pool of items and responses was created (N = 91). The number of items created was smaller than the number of coding units because several coding units were impractical or impossible to implement. Attempts were made to include each coding unit when possible and the primary investigator decided which coding units would not be included in the individual FUI items. The criteria considered in the exclusion of coding units were: (1) whether the coding unit could be adapted into an item administered at a non-individual level and (2) whether the coding unit was directly relevant to factual knowledge of the legal system. Examples of coding units excluded based on the first criteria were the defendant's mental status, the details of the defendant's specific crime, the defendant's version of the offense they are charged with, and whether the defendant is willing to accept a plea bargain for their current charges. Examples of coding units excluded based on the second criteria were the defendant's general history, mental health history, and whether the defendant believes he or she has a mental illness. The coding units excluded based on the second criteria also fit into the exclusion category for the first criteria (e.g., it is impractical to ask a defendant his or her mental health history on an instrument designed to be administered on a group level). After a review of the items for grammar, practicality, and relevance, 88 items remained on the instrument for implementation in Study 4 (Appendix A). The next step in construction of the FUI was to send the individual items out to experts in the field for review.

Study 4

Method

The 88 items constructed for the instrument were sent out to expert reviewers to receive input on the items as a means of targeting problematic items and item responses before administration of the instrument in the pilot study (Study 5).

Participants. Expert reviewers were 20 mental health professionals with extensive knowledge and experience regarding competency to stand trial evaluations, either through research, practice, or both. Participants in the study were recruited by four separate means. At the time of Study 2, seven out of the ten participants agreed to participate in Study 4. These seven participants were contacted for participation (approximately three months later). Thirty-three additional participants were recruited by either: (1) referrals from the seven mental health professionals from Study 2, (2) they were known to the primary investigator through academic and/or clinical experience, or (3) an American Psychology-Law Listserv announcement (Appendix B).

Demographic characteristics of 19 expert reviewers are reported in table 5 (one participant did not provide demographic information). Fifty-eight percent of the participants were male. The mean age was 44.78 (*SD* =10.75). The majority of participants reported their ethnicity as White/Caucasian (84.2%). All but one participant held a Ph.D., Psy.D., or M.D. (94.7%), and a range of post-doctoral years practicing in the field was reported. Most participants (68.4%) endorsed spending a significant amount of their professional time in forensic work (>75% of the respondents clinical or research work). A range of time devoted specifically to work in competency to stand trial was reported, with the largest amount of participants spending 90-100% (36.8%).

Table 5

Expert Reviewer Demographics (N = 19)

Characteristic	*n*	%
Gender		
Female	8	42.1
Male	11	57.9
Age		
25-35	4	22.2
36-45	4	22.2
46-55	7	38.9
56-65	3	16.7
Ethnicity		
White/Caucasian	16	84.2
Asian/Pacific Islander	1	5.3
Multi-Racial/Multi-Ethnic	1	5.3
Other	1	5.3
Highest level of education		
Ph.D.	15	78.9
Psy.D.	2	10.5
M.D.	1	5.3
M.A./M.S.	1	5.3
Setting primarily worked in		
Forensic/Prison	7	36.8
Private Practice	4	21.1
Inpatient Hospital	4	21.1
University/Research/Academic	3	15.8
Other	1	5.3
Post-doctoral years practiced in the field		
1-5	5	27.8
6-10	7	38.9
11-15	2	11.1
16-20	2	11.1
21-25	2	11.1
How much of clinical work/research is devoted to forensic work (in percentage)		
<50%	2	10.5
51-75%	4	21.1
76-100%	13	68.4
How much of clinical work/research is devoted to competency to stand trial (in percentage)		
10-29%	6	31.6
30-49%	1	5.3

50-69%	3	15.8
70-89%	2	10.5
90-100%	7	36.8

Procedure. Upon agreement to participate in the study all participants were provided with an informed consent document, demographics form, instructions, and expert review forms (Instructions and expert review form example are in Appendix C). Materials were sent out to the expert reviewers either through electronic mail or postal mail. The instrument items were divided in half. Each expert reviewer received 44 items. Items were divided such that half of the reviewers (n = 19) rated the first half of the items (item set A) and the remaining reviewers (n = 21) rated the latter half of items (item set B). The reviewers were asked to rate each individual item on the difficulty level, grammar/wording of the question and response options, accuracy of correct response, plausibility of incorrect response, and the overall quality of the item on a 5-point Likert scale. Respondents were asked to provide qualitative feedback on items that received a rating of less than "acceptable" and were asked to provide any feedback to improve the item. It was hypothesized that the expert reviewers would rate the subtle items as significantly more difficult than the obvious items and that a range of evaluative comments would be provided.

Results and Discussion

Of the 40 participants recruited, 20 returned the expert review forms (50% return rate). Independent *t*-tests were conducted to test the hypothesis that the items with an obvious answer choice would be rated as less difficult than the items with a subtle answer choice. On item set A obvious items were rated as significantly less difficult than subtle (t = -2.85, df = 14, p < .05).

Similar results were obtained from item set B, in that obvious items were rated as significantly less difficult than subtle items (t = -6.20, df = 9.83, p < .001).

Qualitative comments were analyzed and were incorporated into the items as deemed appropriate. For the majority of items, the comments were readily adapted into the item to improve the overall quality of the wording/grammar, accuracy of the correct answer, and plausibility of the incorrect answer. Multiple word changes and suggestions were incorporated. After a thorough review of all comments, 15 items were discarded from the pool. Examples of reasons for discarded items were (1) the item did not clearly relate to competency to stand trial, (2) the item had too many possible interpretations and could not constructively be improved upon/ambiguity in response options, (3) the item was not applicable across all jurisdictions, and (4) the item was too difficult for most defendants and/or information is not necessary for a defendant to know to be considered competent to stand trial.

After evaluation of all expert review data, 73 items remained on the instrument for implementation in Study 5. Thirty-six items contained an obvious response option and 37 contained a subtle response option. The final sample of FUI items that were used in Study 5 is presented in Appendix D.

Study 5

Following the construction and refinement of test items, the FUI was administered to a sample of college undergraduate students. There were three goals for this pilot study. The first goal was to conduct item analyses to determine which items performed best in the sample and best represent the construct being measured, and additionally to compare participant's performance on the instrument to a validated measure of competency to stand trial. The second goal was to examine the effects of response style (honest vs. feigning), item type (obvious vs. subtle), and intelligence level in the prediction of individual's responses to test items. The third goal was to examine the test's performance against the forced-choice testing theory and completion time methodology. Clinically impaired samples will be needed to establish a floor effect, and therefore data to test this detection method were not obtained in the preliminary phase of test development and validation.

Method

Participants. Two hundred and fourteen undergraduate students from John Jay College of Criminal Justice at The City University of New York were recruited from introductory psychology courses for participation. All participants who completed the battery of tests received course credit for their participation.

Demographic variables for the participants are presented in Table 6. Participants were predominately female (67.6%); the mean age was 19.96 (*SD* = 4.39). The majority of participants were freshmen (65.3%).

Table 6

Pilot Study Participant Demographics (N = 214)

Characteristic	*n*	%
Gender		
Female	144	67.3
Male	70	32.7
Age		
18-20	180	84.1
21-23	16	7.5
24-26	7	3.2
26+	11	5.2
Ethnicity		
Hispanic/Latino	84	39.3
African American	39	18.2
White/Caucasian	32	15.0
Asian/Pacific Islander	26	12.1
Multi-Racial/Multi-Ethnic	17	7.9
Other	14	6.5
Native American	2	0.9
Class Standing		
Freshman	137	64.9
Sophomore	51	24.2
Junior	17	8.1
Senior	6	2.8

Materials. All participants in the study completed a battery of three instruments. The test materials consisted of the Shipley Institute of Living Scale (Shipley, 1940), the Factual Understanding Instrument (FUI) constructed for this study, and the CAST-MR (Everington & Luckasson, 1992). The SILS was administered as a measure of general intellectual functioning and the CAST-MR was employed to serve as a measure of convergent validity for the construct of competence-related knowledge.

Shipley Institute of Living Scale (SILS). The SILS (Shipley, 1940; Zachary, 1986) is a test of general intellectual functioning appropriate for adolescents and adults aged 16-64. The SILS consists of two subtests, Vocabulary and Abstractions. The Vocabulary subtest consists of 40 multiple-choice items in which the examinee is to choose which of the four presented words corresponds closest to the target word. The Abstraction subtest consists of 20 questions in which the examinee is to choose the correct character to complete the pattern of a string of numbers, letters, or words. Both subtests are timed; participants are allowed 10 minutes for each subtest. An individual score for the Vocabulary and Abstraction subtests is calculated for each examinee and a total score (summation of both subtest scores) is then computed.

Normative data for the SILS was collected from populations of high school, college, and nursing students; psychiatric patients; and hospital employees. Research suggests the reliability of the instrument is adequate to good, with total score test-retest coefficients ranging from .60-.82 and an internal consistency score of .92. Additionally, the SILS has empirical support as a valid estimator of current intellectual functioning, with correlations between the SILS total score and the WAIS-R full scale IQ score reported between .74-.85. The SILS total *T*-score was used as to operationalize intelligence in the current study.

Factual Understanding Instrument (FUI). The FUI consists of 73 items each with two possible answer choices. The instrument is completed individually where each person is asked to read each item and circle one answer. Subscale scores are calculated for the number of correct subtle items, the number of correct obvious items, and a total score is calculated by summing the two item type scores. Completion time (in minutes and seconds) is recorded as the time it takes the participant to answer all the items on the FUI.

Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR). The CAST-MR (Everington & Luckasson, 1992), an instrument used in the assessment of competency to stand trial for individuals with mental retardation, utilizes a multiple-choice and open-ended format. The instrument includes 40 multiple-choice questions followed by 10 open-ended questions specific to the defendant's case, all written at a 4th grade reading level. The areas assessed are basic legal concepts (section I, 25 questions), skills to assist defense (section II, 15 questions), and understanding of case events (section III, 10 questions). Section III was not administered to participant's in this study because this section is specific to a defendant's current legal charges and thus would not be applicable to students who are not actually facing criminal charges. A score of 0-40 (sections I and II) was used as the outcome variable in the current study.

Research has supported the reliability and validity of the CAST-MR as a measure of competence to stand trial knowledge (Everington & Luckasson, 1992). Initial validation studies utilizing a known-groups design support the criterion validity of the instrument with significantly higher scores produced by those deemed competent vs. incompetent to stand trial with an 83-100% correct classification rate for competent and incompetent defendants based on test scores (Everington, 1990; Everington & Dunn, 1995). High internal consistency scores (alpha = .92) and a test-retest reliability of .90 have also been found for the CAST-MR when all three sections are administered. Although the authors of the CAST-MR have used the instrument with the omission of section III for similar reasons as offered above (Everington et al., 2007), the psychometric properties of the modified instrument have not been investigated.

The standard administration of the CAST-MR is through oral presentation of the items; however, it was adapted for the current study to be self-administered whereby the participant

read each item on their own and circled an answer. It was hypothesized that self-administration of the CAST-MR would minimally impact the psychometric properties and interpretation of the instrument because oral presentation of the CAST-MR is required because the test is intended for persons of low cognitive ability who generally have limited reading skills. The student sample utilized in this study was not expected to have significant reading deficiencies.

Procedure. Participants were run in a group format of 5-10 persons per session. The primary investigator and a trained M.A. level research assistant ran all sessions. Four experimental conditions were implemented. Two instructional manipulations were used and within each manipulation, the CAST-MR and the FUI were presented in a counterbalanced order. The experimental condition was randomly applied to each session. Sessions were run so that all participants in a given group were either in the honest condition or the feign condition.

All testing materials for each participant (consent form, demographics form, SILS, instructions, FUI, CAST-MR, manipulation check, and debriefing form) were included in a numbered packet each participant received upon entering the room. After completion of the consent form and demographics form, participants were instructed to complete the SILS. Instructions were then read aloud by the experimenter and participants were instructed to read along. Participants were allotted 10 minutes to complete the Vocabulary subtest and 10 minutes to complete the Abstraction subtest.

Participants were then instructed as to how to respond to the FUI and the CAST-MR. Two sets of instructions were utilized; one for the honest group and one for the feign group (Appendix E). Each experimental session received only one set of instructions. Detailed written instructions were provided to each participant while the experimenter read them aloud. Participants in the honest condition were told that they would be taking two tests that are often

given to people who are going to court for a criminal trial to see what they know about the legal system. They were instructed to perform to the best of their ability on the tests. Participants in the feign condition were also told that they would be taking two tests that are often given to people who are going to court for a criminal trial to see what they know about the legal system. However, they were told that they should answer the questions as if they had a lack of knowledge of the legal system and how it works. They were instructed to try to appear to lack legal knowledge but not let on to the tests that they were "faking."

After the participants were instructed, and all questions regarding their performance were answered, the participants were asked to complete the first legal related knowledge instrument (either the FUI or the CAST-MR). Participants were asked to circle only one answer for each of the 73 items on the FUI and only one of the answers for each of the 40 items on the CAST-MR. Participants were instructed to start the first instrument at the same time. The experimenter recorded the time each participant began the instrument using an unobtrusive digital timer. Participants were asked to raise their hand after completing the first instrument and as each participant finished the first instrument, the time was recorded. Participants were also asked to write down the time in which they began and completed the instrument. This was requested as a back-up to the experimenter's completion time recording. All participants were asked to turn the first instrument upside down and to sit quietly until everyone in the group had completed their first instrument. After all participants in the session completed the first instrument, they were asked to take out the second legal related knowledge instrument in the packet (either FUI or CAST-MR). Participants were reminded of the instructions (i.e., honest or feign) and all participants began the instrument at the same time. Again, participants were instructed to raise their hand upon completion of the instrument and each individual's completion time was

recorded. Participants were also instructed to enter the time the instrument was started and completed on the form.

In all conditions, after the last instrument was completed, each participant was asked to fill out a manipulation check. The questions included: how well they understood the instructions they received, how well they felt they were able to follow the instructions, and whether they put effort into their performance on the tests. All responses were on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). In addition, participants in the feigning condition were asked whether they believed the tests would be able to detect whether or not they were feigning. Participants were provided with a debriefing form that described the purposes of the research.

Results and Discussion

Manipulation check. Prior to conducting item analyses, responses to the manipulation check were evaluated. Participants reported strong agreement with the following statements: that they understood the instructions (M = 4.72, SD = .48), were able to follow the instructions (M = 4.71, SD = .48), and put effort into either answering honestly or feigning their performance on the instruments (M = 4.58, SD = .57). The majority of the participants in the feign condition reported not knowing whether the instruments would be able to detect they were feigning (60.2%). Approximately one quarter of participants (25.6%) believed the instruments would be able to detect feigning, and 14.2% believed the instruments would not.

FUI item analysis. For the purposes of FUI item analysis, results are reported for the honest participants (n =101), the feign participants (n = 113), and the combined sample (N = 214). In Appendix F the means, standard deviations, and point biserial correlations for each item within each sample are listed (Table F1). In the combined sample, with the exception of one item (FUI2), all items were answered correctly by at least half of the participants. Mean response

(from 0-1) to FUI items in the combined sample was .75 (*SD* = .11; range .45-.96); on average 75% of the participants answered an item correctly. In the honest condition all items were answered correctly by at least 72% of the participants with the exception of one item (FUI2). Furthermore, within the honest condition, the majority of items were answered correctly by at least 90% of the sample (mean response to the FUI items was .93 [*SD* = .07; range .67-1.00]), indicative of a ceiling effect. In the feign condition, several items were answered correctly by less than half of the participants. Mean response to the FUI items in the feign condition was .58 (*SD* = .18; range .25-.93).

Figure 1 displays the cluster of scores on the FUI for the combined sample. There are two groups, one with scores clustered around 65-73 and another group with scores clustered around 30-50.

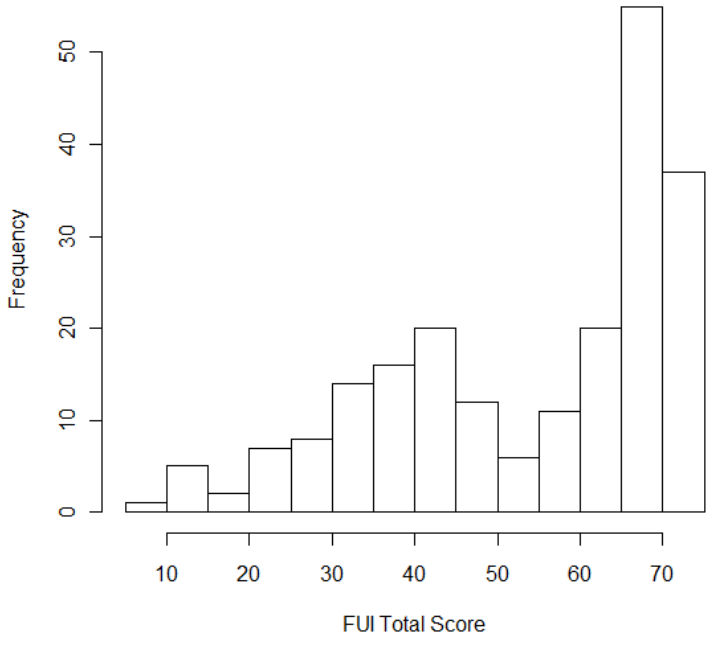

Figure 1. Frequency of Factual Understanding Instrument (FUI) scores for the combined sample (*N* = 214). Total score is calculated out of 73.

When separated by response condition, FUI scores for the honest responders clustered around 65-73 (Figure 2); however in the feign condition, responses were more variable and appeared to be normally distributed (Figure 3).

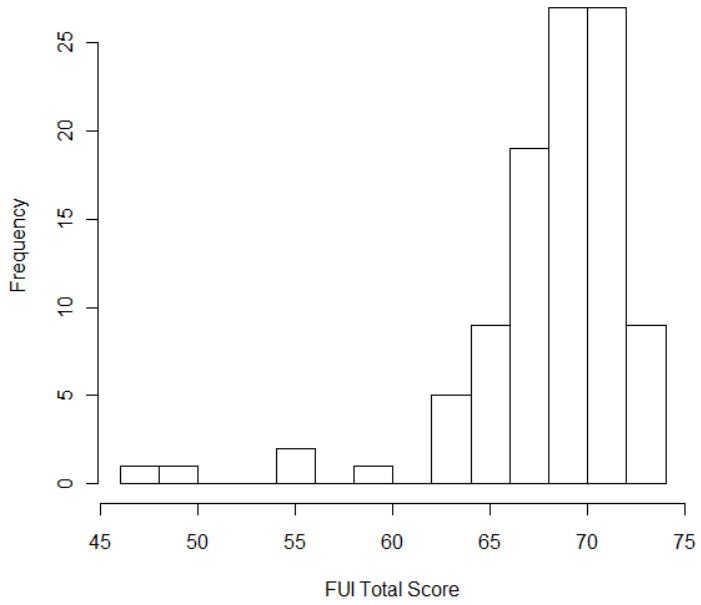

Figure 2. Frequency of Factual Understanding Instrument (FUI) scores for the honest sample (*n* = 101). Total score is calculated out of 73.

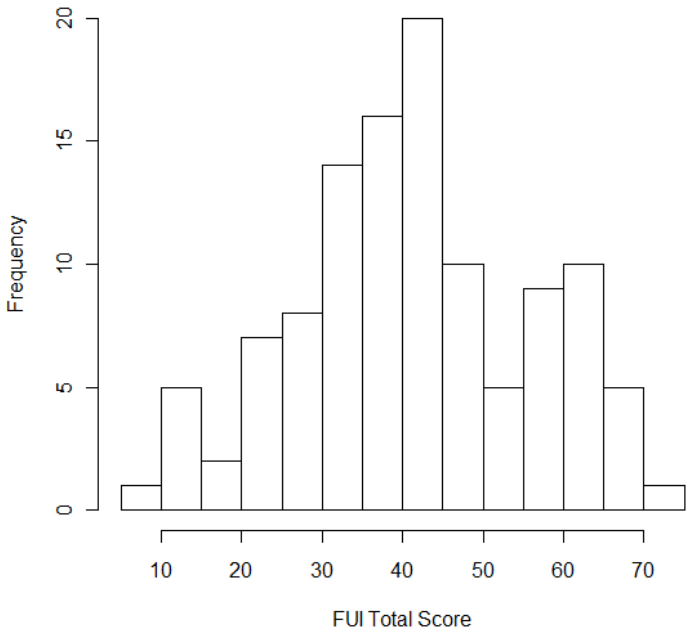

Figure 3. Frequency of Factual Understanding Instrument (FUI) scores for the feign sample (*n* = 113). Total score is calculated out of 73.

Due to a ceiling effect within the honest condition (i.e., every participant answered the item correctly), 10 of the FUI items were removed from further analysis due to zero variance surrounding the item. These 10 items were removed for the honest sample reliability analysis, IRT analysis, and factor analysis (FUI1, FUI7, FUI8, FUI10, FUI 26, FUI30, FUI35, FUI37, FUI51, and FUI52).

Measures of internal consistency (Cronbach's alpha) for the combined, honest, and feign samples were .97, .82, and .94, respectively. Guttman's lamba 2 statistic, an alternative test of reliability which utilizes item covariance, was also estimated for the honest sample (lambda 2 = .83). Spearman-Brown split-half reliability coefficients were calculated. For the combined sample the FUI halves were highly correlated (*r* = .96). Results were similar for the feign sample (*r* = .92), but reliability was slightly lower in the honest sample (*r* = .81). In the combined

sample, the correlation between each item and the remainder of the FUI items (point biserial correlation) ranged from .28-.77; in the honest condition correlations ranged from .00-.70; and in the feigning -.04-.58. Standard errors were approximated for the point biserial correlations using the formula $1/(N-1)^{1/2}$ (Crocker & Algina, 1986). The critical point biserial value was set at two standard errors above .00 for the retention of FUI items. Critical values for the combined, honest, and feign samples were .14, .20, and .19, respectively.

Based on the item analysis, several conclusions can be reached. For honest responding participants FUI items were relatively easy to answer correctly. The exception was in two items, one a subtle item which sampled knowledge of what happens at an arraignment (FUI2), and the other, a subtle item, which questioned where people go after being found not guilty by reason of insanity (FUI61). A potential reason for the difficulty with FUI2 was that the incorrect answer was quite similar to the correct answer. Another possibility is that lay persons typically do not have the knowledge sampled by these two questions, particularly in an area as detailed as not guilty by reason of insanity.

When participants were asked to feign a lack of knowledge of the legal system, 28/73 items were answered correctly by less than half of the participants. The remainder of the FUI items had varying proportions of participants answering correctly. This suggests that when participants were actively suppressing their performance, FUI scores were spread across the range, with an overall chance of correctly answering an item varying from below what would be expected by symptom validity testing (i.e., less than 50%) to ranges that were seen within the honest responding participants.

In the three samples (combined, honest, and feign) FUI reliability analyses were high. Interestingly, the honest sample had the lowest internal consistency and split-half values. This

may be an artifact of the removal of 10 items with zero variance, as longer tests tend to have higher reliability estimates (the majority of the items that were answered correctly by all participants were in the first half of the FUI). In addition, within the honest condition several items did not correlate highly with the overall FUI score, a result of low item variance restricting the point biserial values. When response conditions were combined, FUI items correlated at least .30 with the total score, suggesting that answers to individual items were related to the performance on the entire FUI. In the feign condition, five items did not correlate highly with the FUI total score even though there was variance surrounding the item (i.e., values below the .20 critical value). Although point biserial values were low for five items in the feign sample, these items (i.e., FUI2, FUI4, FUI16, FUI34, and FUI58) performed adequately within the combined and/or honest sample and therefore are not good candidates for elimination from the instrument in this preliminary stage of test construction and analysis. In addition to this reasoning, another potential reason for retaining multiple items that tap into the same knowledge (e.g., several items question the job of the defense attorney) is for analysis of consistency in item endorsement. That is, these repeat items could be used to answer the question of whether feigning participants answer consistently across items measuring the same knowledge. This empirical question was not investigated during the pilot study, but would be of interest in future investigations.

FUI IRT analysis. As a further exploratory analysis into the performance of the FUI items within the response conditions, item difficulty (ability level on the latent trait of knowledge of the legal system at which 50% of the examinees answered an item correctly) and item discrimination (how well the item distinguished between participants with a range of abilities on the latent trait) parameters were estimated in the *R ltm* package for latent traits models under Item Response Theory (R Development Core Team, 2008). Two models were run for each

sample to estimate the difficulty and discrimination parameters (rasch and ltm). The rasch model, a one parameter logistic model assumed all discrimination parameters were equal. The ltm model, a two parameter logistic model assumed a different discrimination parameter per item (Rizopoulos, 2006). The two models were compared through a likelihood ratio analysis to determine if the two parameter model was necessary. The likelihood ratio was significant for the combined sample (LR = 412.58, $df = 72$, $p < .001$), the honest sample (LR = 137.84, $df = 62$, $p < .001$), and the feign sample (LR = 238.88, $df = 72$, $p < .001$). The discrimination parameters significantly differed from one and therefore the ltm model best described the data for all three samples. The difficulty and discrimination parameters for each item are reported in Appendix F (Table F2).

Difficulty levels ranged from negative infinity to positive infinity, with a negative value reflecting an easier item and a positive number reflecting a harder item. Discrimination values ranged from zero to infinity, with a value of zero denoting the item as providing no discrimination value. The relationship between point biserial correlations from the item analysis and discrimination values from item response theory was apparent; as correlations increased, discrimination also improved.

For the combined sample, examples of easier items with low difficulty values were FUI39, FUI1, FUI40, FUI52, and FUI70. Discrimination parameters varied throughout the items (.62 - 3.77). For example, FUI5 discrimination equaled .62, suggesting that the item did not differentiate well between participants with a range of ability on the latent trait of factual knowledge of the legal system; whereas FUI10 has a discrimination value of 3.77, suggesting the item performed well in distinguishing between participants of varying ability levels. Figure 4 depicts the relationship between item difficulty and discrimination for the combined sample. As

shown, the difficulty and discrimination values were variable, with the majority of items having difficulty values below zero and discrimination values between +1 and +3.

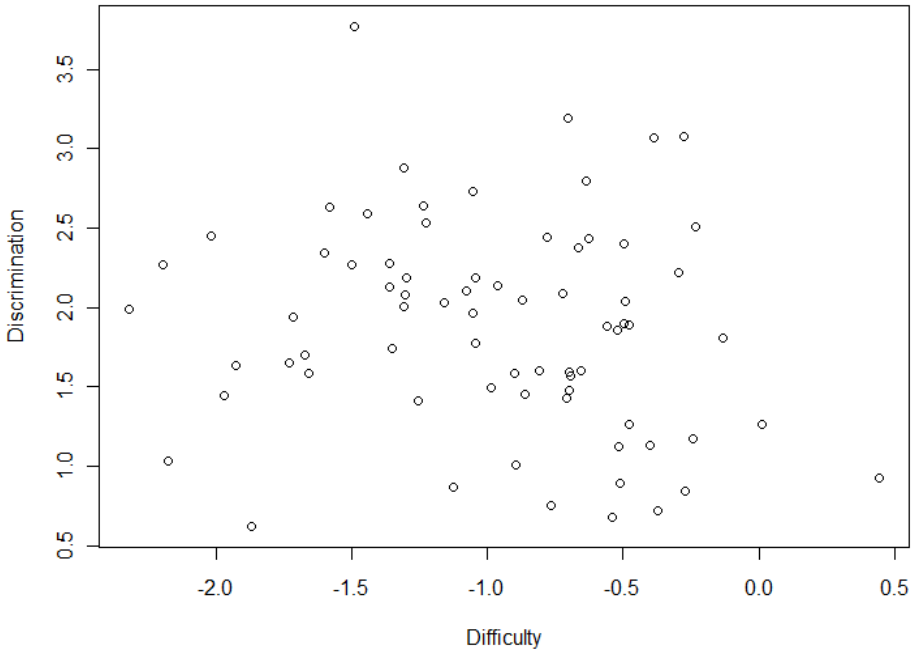

Figure 4. Relationship between Factual Understanding Instrument (FUI) item difficulty and item discrimination in the combined sample (N = 214).

After the difficulty and discrimination parameters were estimated for the combined sample, Welch two sample *t*-tests were run to compare the difficulty and discrimination values between the a priori defined subtle and obvious items as it was hypothesized that the parameters would differ. Difficulty level did not significantly vary between subtle and obvious items (t = .17, p = .86, *ns*), nor did discrimination (t = .46, p = .65, *ns*).

In the honest sample, examples of items with low difficulty values were FUI65, FUI62, FUI43, FUI23, and FUI6; the most difficult items within this sample were FUI24 and FUI40. FUI29 was an item with a high discrimination value, suggesting that it differentiated well

between responders with varying ability levels. As shown in Figure 5, the majority of the FUI items had low difficulty values, ranging from -5 to +5 and had discrimination values ranging from 0 to +3.

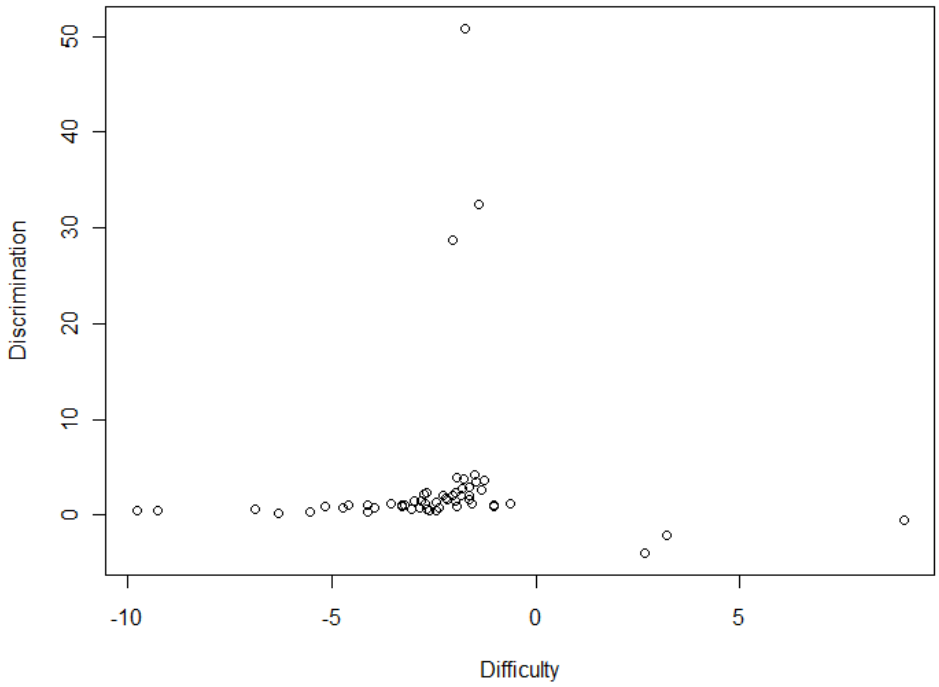

Figure 5. Relationship between Factual Understanding Instrument (FUI) item difficulty and item discrimination in honest sample (n = 101). FUI items 24, 40, 62, and 65 were deleted from the figure for ease of interpretation. All had outlier difficulty values.

Welch two sample *t*-tests comparing the difficulty and discrimination parameters between the a priori defined subtle and obvious items for the honest condition did not significantly differ (t = -1.16, p = .25, *ns*; t = 1.54, p = .13, *ns*, respectively).

In the feign sample, examples of items with a low difficulty value were FUI1 and FUI39; examples of items with high difficulty values were FUI2 and FUI16. Items with high discrimination values were FUI39 and FUI49; items with low discrimination values were FUI2

and FUI16. As shown in Figure 6, the majority of FUI items in the feign sample had difficulty values ranging from -2 to +2 and had discrimination values ranging from 0 to +3.

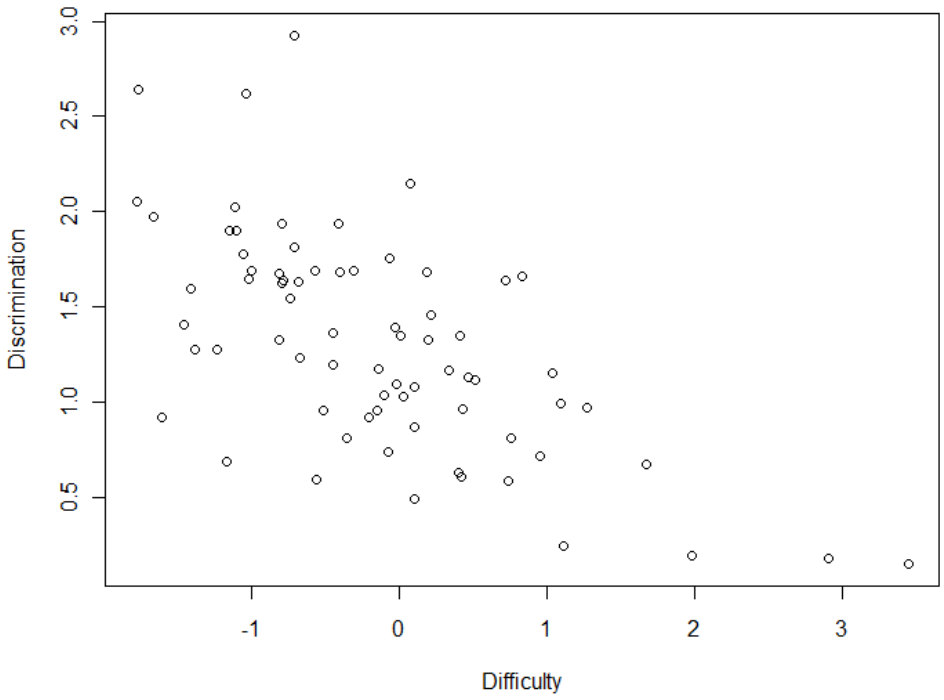

Figure 6. Relationship between Factual Understanding Instrument (FUI) item difficulty and item discrimination in feign sample (*n* = 113). FUI item 12 not included in this graph for ease of interpretation; FUI12 had an outlier difficulty value.

Similar to the combined and honest conditions, difficulty and discrimination parameters did not significantly differ between item type in the feign sample (*t* = .65, *p* = .52, *ns; t* = .07, *p* = .941, *ns*, respectively).

From the analysis of the FUI difficulty and discrimination values, several conclusions can be reached. First, within the honest sample, all difficulty values were in the negative range (with the exception of four items) reflecting their level of ease. However some FUI items performed better in distinguishing between participants with varying levels of knowledge. For example,

FUI65 had the lowest difficulty value and in turn, the discrimination value of this item was very low (almost zero). Because the item did not require a lot of knowledge to answer correctly, it was not useful in distinguishing between participants with varying ability levels. Other items did appear to differentiate well between participants and these items would be of interest to investigate further in other populations.

Second, although the FUI items had low difficulty values within the honest sample, responses to the FUI items by honest participants produced the greatest range of difficulty values. This is consistent with the population as responding with effort, hence the underlying trait was more closely approximated than it was in the feign sample. The difficulty and discrimination parameters were more difficult to interpret in the feign sample, as the latent trait underlying the performance by these participants was altered and does not reflect their true abilities. However, some preliminary statements can be made about the feign sample. A few items (FUI2, FUI16, and FUI58) appear to be items that participants consistently got incorrect. In fact, throughout item analysis and IRT analysis, FUI2 was the most difficult item for all participants. Similarly, there were a couple items that feign participants tended to answer correctly, perhaps due to their level of ease on face value. It is of interest that the items that were generally answered correctly by feign participants were the same items that were answered correctly by the majority of the honest participants. For example, FUI1 was answered correctly by all honest responders and this was the item with the lowest difficulty value in the feign sample. This was a consistent trend in the analyses; items answered correctly by all (or most) honest participants were also answered correctly by a high percentage of feign participants.

Third, contrary to the hypothesis, difficulty and discrimination parameters did not significantly differ between the subtle and obvious items in each response condition. Subtle

items did not require more knowledge to correctly answer than did the obvious items. Based on all item analyses, it appears that the majority of items on the FUI were relatively easy for participants and therefore items that were defined as more difficult, were not in fact difficult for the two groups of college undergraduates.

Factor Analysis. Exploratory Factor Analysis was conducted on the FUI in each of the three samples. Although exploratory, it was hypothesized that items that tapped into the same information would load on the same factor (e.g. all items questioning the job of the prosecutor would load on one factor). Prior to running the analyses, all data were converted to polychoric correlations (as FUI items have a dichotomous outcome value) using the R *polycor* package. The correlation matrices were analyzed for values above .3 to ensure that factor analysis was an appropriate analysis for the data. The majority of FUI items were correlated at or above this value.

Due to the ceiling effect and the high correlations between the FUI items within the honest condition, the correlation matrix could not be inverted, and no factor structure could be computed. Therefore, factor matrices were calculated only for the combined and feign samples. Principal components extraction was applied and several nonrotated and rotated solutions were computed (i.e., one, two, three, four, and five factor solutions); orthogonal varimax and oblique promax rotations were applied. As FUI factors were highly correlated (combined sample $r = .64$, feign sample $r = .49$), oblique promax was the appropriate rotation.

For both the combined and feign samples, a two factor solution was the best fit. In the combined sample, the two factor solution accounted for 58.7% of the variance (factor 1 σ^2 = 29.7%, factor 2 σ^2 = 29%). In the feign sample, the two factor solution accounted for 42% of the variance (factor one σ^2 = 23.5%, factor two σ^2 = 18.5%). Item loadings and communalities for

both samples are presented in Appendix F (Table F3). A cutoff of .30 for factor loadings was applied, and values less than this are not reported. In the combined sample, factor loadings and communality values tended to be high, and thus none were excluded. However, 10 items loaded highly on both factors. In the feign sample, three items (FUI4, FUI16, and FUI36) had loadings of less than .30 and were excluded. In the feign sample, communalities were not as high as in the combined sample, reflecting the heterogeneity of responses among this sample of participants. This may also be a result of the small sample in the feign condition (a minimum sample size of 200 is typically recommended for these analyses). Two items had similar loadings on both factors in the feign sample.

Although not exact, a similar pattern of item loadings was found across response conditions (i.e., items which loaded on factor 1 in the combined sample also tended to load on factor 1 in the feign sample). Contrary to what was hypothesized, the FUI items that loaded on the factors within each sample were not theoretically connected. Each factor included items that tapped into the courtroom players, sentences and pleas, and courtroom procedures/terminology. These three domains appeared in each of the two factors in each sample. For example, in the combined sample, factor 1 included items which tapped into knowledge of the job of the judge, jury, prosecutor, and defense attorney. Factor 2 included items which tapped into knowledge of the job of the defendant, witness, defense attorney, and judge. It appeared that items of the same content (e.g., job of the defense attorney) did not load on the same factor. This pattern also occurred in the feign sample. For example, factor 1 included several items surrounding the job of the jury, judge, and prosecutor and factor 2 included items which tapped into knowledge of the job of the prosecutor, defense attorney, and witness.

These factors were not easily interpreted on face value. However, what is of interest is that within both samples, factor 1 predominantly included subtle items and factor 2 predominantly included obvious items. For example, in the combined sample, factor 1 consisted of 30 subtle items and five obvious items and factor 2 contained 25 obvious items and three subtle items. From this first stage of FUI analysis, it can be surmised that the two factors which made up the FUI in the student sample represented the split between the item types of subtle and obvious.

Comparative analyses of FUI and CAST-MR scores. To test the hypothesis that those who were feigning would produce lower scores on FUI items, independent sample *t*-tests were run to compare the scores from each response condition (honest and feign). Significant differences were found between response condition on the FUI total score ($t = 17.55$, $df = 212$, $p < .001$), the FUI obvious items score ($t = 12.64$, $df = 212$, $p < .001$), and the FUI subtle items score ($t = 19.53$, $df = 212$, $p < .001$). Means and standard deviations for the two response conditions are reported in Table 7.

Table 7

Scores on the Factual Understanding Instrument (FUI) by Response Condition and Item Type

	Total		Obvious		Subtle	
Condition	*M (SD)*	*t*-value	*M (SD)*	*t*-value	*M (SD)*	*t*-value
Honest	68.51 (4.40)	17.55*	35.14 (1.56)	12.64*	33.38 (3.25)	19.53*
Feign	41.94 (14.63)		24.58 (8.27)		17.31 (7.67)	

Note. Total score is computed out of a total of 73; Obvious score is computed out of a total of 36; Subtle score is computed out of a total of 37.
*significant at $p < .001$

It was also hypothesized that participants in the feign condition would produce lower scores on the CAST-MR than those in the honest condition. Participants in the honest condition produced a mean score of 37.54 ($SD = 3.04$). Feigning participants had lower scores on the CAST-MR ($M = 19.43$, $SD = 9.23$) and this difference was statistically significant ($t = 18.82$, $df = 212$, $p < .001$).

These comparative analyses provide further support for the notion that students were able to feign their performance or respond honestly on the FUI and CAST-MR and, in addition, that their performance was consistent across the instruments. As hypothesized, the FUI total score was significantly lower in the feign sample; however this score was above the 50% mark that would follow from symptom validity testing. As previously noted this method has not been sensitive enough as a feigning detection method in other instruments and this notion held true for the FUI. It appeared that participants feigned their performance, but not in such an extreme manner as to be overly obvious. It is also of interest that in both response conditions, participants produced higher scores on the obvious items than the subtle items. This finding supports the notion that the obvious items were easier to answer correctly.

Comparative analyses of completion time. To test the hypothesis that those who were feigning would take a greater amount of time to complete the CAST-MR and FUI, independent sample *t*-tests were run to compare the completion times of participants in the feign and honest groups on these instruments. Prior to conducting *t*-tests, the data were analyzed to determine if a single response time population existed within each instrument and response condition by

running a latent mixture distribution analysis for univariate data. The *R mixdist* package was used for this analysis (Macdonald & Green, 1988).

When the parameters of the completion time distribution were estimated, the two population model fit best for the honest conditions on both the FUI and CAST-MR (i.e., there were two populations of response times within each instrument in the honest condition). On the FUI, in the honest responding condition, the two populations were centered around mean response times of 11.32 minutes (*SD* = 2.08) and 14.32 minutes (*SD* = 3.91). On the CAST-MR, in the honest condition, the two populations centered around mean response times of 6.30 minutes (*SD* = .94) and 8.58 minutes (*SD* = 2.10). The two population model also fit best for the CAST-MR response time in the feign condition, with the two populations producing mean response times of 7.84 minutes (*SD* = .40) and 9.36 minutes (*SD* = 2.97). The one population model fit the data as well as the two population model for the FUI response time in the feign condition, therefore, the one population model is the most parsimonious explanation of the completion time data for this sample (*M* = 13.37 minutes, *SD* = 3.46). Figure 7 presents the response time mixture distributions of the four samples (FUI honest, FUI feign, CAST-MR honest, and CAST-MR feign).

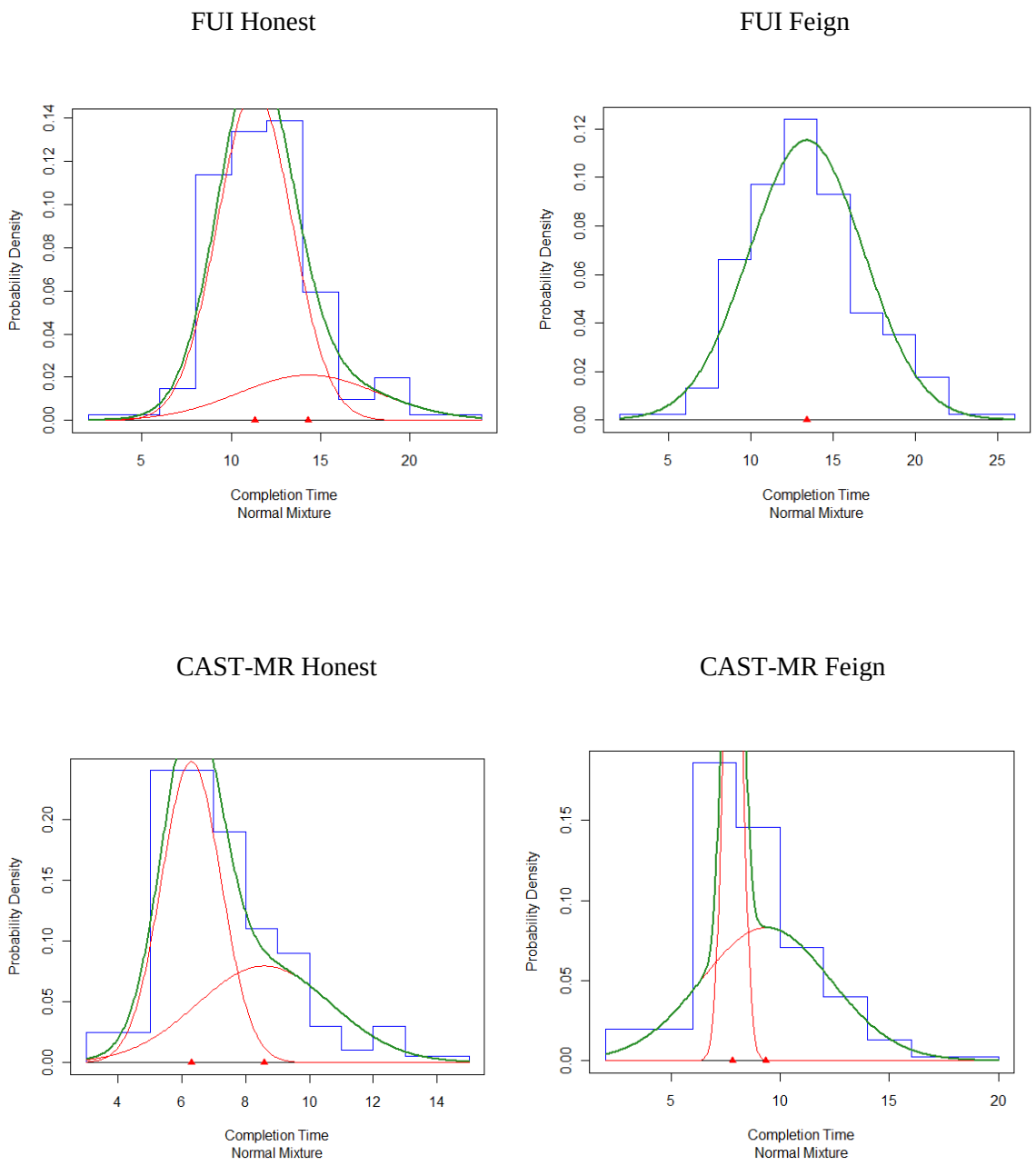

Figure 7. Mixture distributions for FUI (Factual Understanding Instrument) and CAST-MR (Competence Assessment to Stand Trial for Defendants with Mental Retardation) completion time (in minutes) by repsonse condition. Red curves represent individual response populations; green curve represents the combination of the two populations (or one population as in the FUI feign response condition). Probability density is the likelihood of a variable occuring at a given value. *n* =113 for FUI and CAST-MR feign repsonse conditions; *n* = 101 for FUI and CAST-MR honest repsonse conditions.

For the CAST-MR, mean response times and standard deviations were compared between response conditions and populations. The CAST-MR honest population 1 statistics were compared to the CAST-MR feign population 1 statistics (t = 9.18, df = 1, p = .07, *ns*) and the CAST-MR honest population 2 statistics were compared to the CAST-MR feign population 2 statistics (t = 22.99, df = 1, $p < .05$). No differences were found in completion time within population 1 of the CAST-MR, but differences were found in the population 2 with feigning participants producing longer completion times. These findings provide some preliminary support for the use of completion time as a measure of feigning on the CAST-MR. However, if completion time were indeed a good measure of feigning, we would expect to see differences in both response time populations.

A similar procedure was followed for the FUI completion time data, but since only one response time population existed for the FUI feign condition, the two FUI honest population statistics were compared to the one population feign statistics. The statistical comparison of the FUI honest population 1 statistics to the feign population 1 statistics was significant (t = 12.04, df =1, $p < .05$) as was the statistical comparison of the population 2 statistics to the FUI feign population 1 statistics (t = 29.15, df = 1, $p < .05$).

On the FUI, honest and feign participants did have significantly different response times based on response condition. However, for population 1, the difference was in the opposite direction as hypothesized. Completion time was shorter in the feign condition than in the honest condition, contrary to the completion time results found in previous studies on feigning. This finding is especially striking because close to 80% of the honest responders fell within the population 1 statistics, suggesting that participants did not take longer to complete the FUI when feigning, but rather completed it faster. Even with this finding, it is important to note that the FUI

population 2 statistics did differ by response condition, indicating some possible support for the hypothesis. However, as noted, only 20% of the feign participants fell within population 2.

There appeared to have been a small group who took significantly longer to complete the FUI when feigning, but overall, participants completed the instrument faster when feigning. This could be due to insufficient attention to the items or random responding. Another possibility is that the participant's strategy was to get through the items quickly as to not appear aberrant in their completion of the instrument. A future hypothesis for exploration could be that, when feigning, participants are cognizant of completion time and want to finish the instrument promptly as to avoid suspicion from the examiner. An additional consideration is that in this study participants were aware that their completion time was being recorded. Had they not been informed of this, the results may have differed.

FUI and CAST-MR correlations. To assess for the convergent validity of the FUI with a validated measure of competency to stand trial knowledge, Pearson correlation coefficients were computed between participant's total score on the FUI and total score on the CAST-MR. Significant correlations were found in the combined (r = .885, p < .001), honest (r = .544, p < .001), and feign samples (r = .721, p < .001), suggesting that the FUI measured similar content as the CAST-MR.

Mixed effects models. As it was hypothesized that a relationship existed between item type (subtle vs. obvious), response condition (honest vs. feign), and intelligence level (SILS total score) in the prediction of FUI responses, mixed effect models were used to investigate the relationships. The SILS Total *T*-Score mean for the combined sample was 51.53 (*SD* = .64; range 32-66); the Vocabulary *T*-score mean was 46.80 (*SD* = 7.40, range 22-65) and the Abstractions mean *T*-score was 54.17 (*SD* = 6.37, range 32-66). All of the mean values fell within the average

range of intelligence; however a span of values were obtained in this sample, spanning +/- 2 standard deviations.

Four models were tested to determine how variables measured at the person level (response condition and intelligence level) affected relationships occurring at the item level (correct/incorrect responses to subtle and obvious items). The responses to each FUI item were the outcome variables (N = 15,612). The *R lme4* package for mixed-effects models was used for this analysis.

Model 1 tested the hypothesis that item type predicted responses to FUI items and that the effect of item type was dependent upon response condition and intelligence level. Model 2 tested the hypothesis that responses to FUI items depended upon item type and that the effect of item type was dependent upon response condition, but not intelligence level. Model 3 tested the hypothesis that responses to FUI items were predicted by response condition and intelligence, with no effect of item type. Lastly, model 4 tested the hypothesis that only response condition predicted answers to the FUI items and no effect existed for item type or intelligence level.

The four models were compared for statistical significance. Model 2 fit significantly better than model 3 or 4 (χ^2 = 660.40, df= 2, p < .001) and Model 1 fit better than Model 2 (χ^2 = 32.49, df = 2, p < .001). Table 8 includes the random and fixed effects for all the tested models.

Table 8

Mixed Effects Models in the Prediction of Factual Understanding Instrument (FUI) Item Responses

Parameter	Model 1	Model 2	Model 3	Model 4
		Random effects		
Participant/intercept (σ^2)	.82 (.91)	.87 (.93)	.78 (.88)	.83 (.91)
SILS/intercept (σ^2)	.37 (.61)	.05 (.23)	.05 (.22)	
SILS/item type (σ^2)	.15 (.39)			
Condition/intercept (σ^2)	45.52 (6.75)	62.68 (7.92)	1.76 (1.32)	1.76 (1.32)
Condition/item type (σ^2)	1.94 (1.39)	1.86 (1.36)		
		Fixed effects		
Intercept	-4.04 (.33)	-5.26 (.26)	1.67 (.94)	1.70 (.94)

Note. Standard deviations are in parentheses. SILS=Shipley Institute of Living Scale. Item type refers to the subtle or obvious items of the Factual Understanding Instrument (FUI). Condition refers to the honest or feign response condition. *N*=15,612.

Figure 8 depicts the relationship between intelligence level, item type, and response condition in the probability of obtaining a correct response (Model 1).

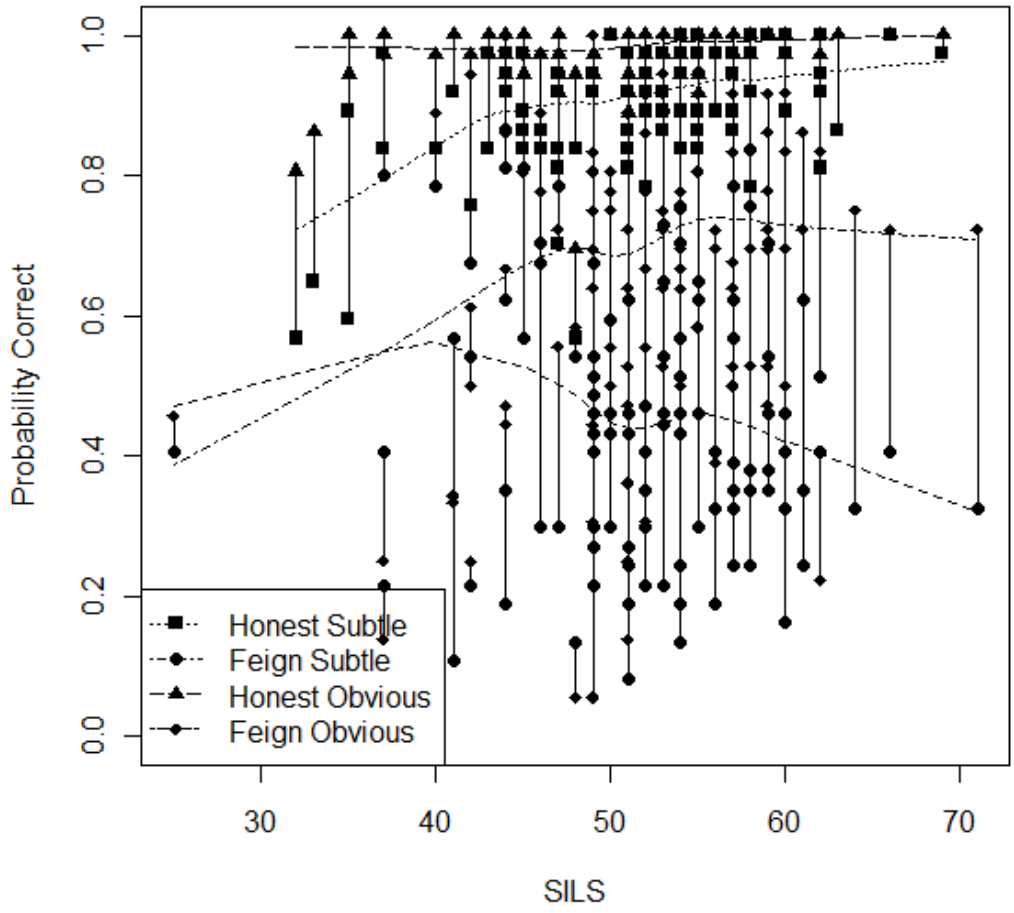

Figure 8. Relationship between intelligence level, response condition, and item type in the prediction of a correct FUI (Factual Understanding Instrument) response. Honest subtle=response to a subtle item by participants in the honest response condition; Feign subtle=response to a subtle item by participants in the feign response condition; Honest obvious=response to an obvious item by participants in the honest response condition; Feign obvious=response to an obvious item by participants in the feign response condition. SILS=Shipley Institute of Living Scale. SILS scores have $M = 50$, $SD = 10$.

In the honest condition, as intelligence increased the probability of correctly answering a subtle item slightly increased; however the probability of correctly answering an obvious item remained high (near 1.0) regardless of intelligence level. This is likely a reflection of the easiness level of the obvious items; nearly all participants could answer them correctly when putting forth

effort, even if intelligence levels fell in the below average range. However, with the subtle items intelligence was a factor in correctly answering a FUI item. The results support the hypothesized relationship between the three variables in the honest response condition.

In the feign condition, a negative relationship between intelligence level and the probability of correctly answering a subtle item was found. As intelligence level increased, participants answered more subtle items incorrectly. Participants with lower intelligence levels (i.e., more than one standard deviation below the mean) had a higher probability of correctly answering a subtle item. Within the obvious items, as intelligence increased, the probability of correctly answering an item increased. Opposite of what was found within the subtle items, participants with lower intelligent levels had a lower probability of correctly answering an obvious item.

This pattern supported some of the original hypotheses about the relationship between the three variables in the feign response condition. As predicted, participants with higher intelligence levels answered more obvious items correctly and more subtle items incorrectly, perhaps as a strategy to appear to not know the answers to the more difficult questions, but to not appear as though obviously feigning. In participants with lower intelligence levels, a decreased probability of getting an obvious item correct was observed, consistent with the hypothesis. This could be a reflection of a haphazard attempt at feigning where one simply answers all questions incorrectly without attending to item difficulty. However, contrary to what was hypothesized, participants with lower intelligence levels had an increased probability of correctly answering a subtle item. It is possible that participants may have not discriminated between item difficulties and therefore randomly chose which items to correctly or incorrectly answer. Another potential explanation is that participants did discriminate between item difficulties, but did not consider this an important

factor when attempting to feign. Lastly, the frequency of participants who fell within the below average intelligence range was quite small, with only nine participants scoring lower than 40 on the SILS. The majority of participants fell within one standard deviation of the mean. This may limit the interpretation of the data. Additional samples with wider ranges of intelligence levels are needed to further explore these relationships.

General Discussion

As a result of the potential consequences associated with proceeding to trial, mental health professionals conducting competency to stand trial evaluations must have a heightened sensitivity to the possibility of feigning in defendants. Because of the high stakes of such evaluations, accurate assessment instruments are needed. The purpose of the current research was to construct an instrument to serve as a screen for a feigned lack of knowledge while simultaneously assessing for factual understanding of the legal system as delineated under *Dusky v. United States* (1960).

Although there are instruments available to assess for feigned mental illness in the context of competency to stand trial, only one attempt has been made at constructing a screen for feigned lack of knowledge and this instrument has several disadvantages to its use. The conceptualization and construction of the FUI, as the first instrument to incorporate multiple feigning assessment strategies into the assessment of competence related knowledge, has attempted to improve upon available instruments. In addition to the incorporation of multiple feigning detection methods, the FUI is also the first instrument of its kind to have a reading level that is more appropriate for forensic populations with a sole focus on the assessment of legal related knowledge. Five studies were conducted on the conceptualization, construction, and evaluation of the FUI, and from these studies several conclusions can be reached on its utility.

First, in regards to the FUI as a measure of legal related knowledge, issues of reliability and validity will be reviewed. The FUI was constructed with a focus on content validity, in that the areas of factual knowledge of the legal system were well represented in the items, both from a literature review and from expert interviews. Expert reviewers provided feedback into the relevancy of each FUI item prior to the use of the instrument in the pilot study, further

supporting the content of the instrument as reflecting the pertinent areas a defendant must factually understand before proceeding to trial. The FUI has also demonstrated adequate construct validity, as the knowledge that is tapped into by the FUI is similar to that of a pre-established measure of competence related abilities. Further explorations of the FUI in conjunction with other pre-established measures of legal related knowledge, administered at an individual level, would further bolster this finding. Lastly, the FUI demonstrated high internal consistency and split-half reliability values, suggesting that the FUI was reliable when administered to the student sample. Additional administrations of the FUI within forensic populations are needed to assess the reliability of the FUI in applied contexts.

Second, the findings from the FUI item analysis indicated that on average, in the honest sample, 93% of respondents correctly answered an item, and in the feign sample, 58% correctly answered an item. Even with these mean item responses, within the feign condition a wide range of scores was found, spanning from less than 50% correct on the FUI to scores seen within the group of honest participants. This finding suggests that when the participants feigned, their performance was quite variable with some appearing more obvious in their strategy and others being more conservative. In item analysis, correlations between the individual items and the overall FUI score were adequate and at this preliminary stage of test construction no items were removed from the instrument. This finding suggests that the items were measuring similar information. Item difficulty and discrimination values estimated in IRT analyses in the honest sample reflected their level of ease. Some items proved to be better at distinguishing between responders with varying levels of knowledge. An interesting trend in these analyses was that the items that participants tended to answer correctly in the feign condition were the same items honest responding participants answered correctly nearly 100% of the time. All of these were

obvious items. The implication from these results is that these particular items are so easy that feigners answered them correctly due to their obvious nature and that furthermore, these items required little knowledge to answer correctly.

It is important to note that the majority of FUI items were generally easy for an honest responding college student sample with no significant cognitive deficits. In fact, a ceiling effect was observed in the honest responding participants, with many scoring near perfect or perfect on the measure. Similarly high scores were found on participant's performance on the CAST-MR. These findings are significant, because as noted, forensic populations generally have less formal education than non-incarcerated populations. If the FUI had been difficult for college students, it would be impractical to attempt to use the instrument with forensic populations. Further explorations of the FUI with forensic populations are necessary to determine the level of item difficulty for participants in these contexts.

Third, the use of subtle and obvious items and response condition in combination with an estimation of intelligence showed promise for the prediction of responses to FUI items, in that depending on intelligence level, participants answered a certain group of items correctly or incorrectly. With the exception of one finding results were consistent with the original hypotheses. Specific findings were that irrespective of intelligence level, within the honest condition, nearly all participants correctly answered the obvious items, but that as intelligence increased, the probability of correctly answering a subtle item increased. Within the feign condition, as intelligence increased, the probability of correctly answering an obvious item increased whereas the probability of correctly answering a subtle item decreased as intelligence rose. It was originally hypothesized that feigning participants with lower intelligence would get nearly all items incorrect regardless of item type; however this was not the case. It is believed

that the feigning participants with higher intelligence levels chose to incorrectly answer the subtle items as a deception strategy. An explanation for the results in the feigning participants with lower intelligence levels is that instead of answering nearly all the FUI items incorrectly, these participants were haphazard in their responses to items or were unable to differentiate between the subtle and obvious items.

Based on the findings from the IRT analyses that item difficulties did not differ between item types, it is important to question whether the a priori defined subtle and obvious items were variable enough in difficulty level to be differentiated from each other. From other analyses conducted, there is support for the notion that the FUI obvious and subtle items were of a different difficulty level. First, expert reviewers rated the subtle items as more difficult than the obvious during Study 4, suggesting that at face value they are of a different difficulty level. Second, the subtle items grouped together in the factor analysis as did the obvious items, suggesting that there is more shared variance amongst items of each difficulty level. Though conflicting results were found in terms of differentiating obvious from subtle items based on difficulty estimates in the IRT analyses, a potential explanation is that the difficulty estimates were based on the point in which 50% of the participants were able to correctly answer an item. In the honest and combined samples, all items, regardless of difficulty level, were correctly answered by at least half of the sample and within the feign sample, fewer than 40% of the items were answered correctly by less than half the sample. The high scores obtained in the samples may have restrained the estimates and the analysis may not have been sensitive enough to detect differences in the items.

As the primary purpose of the FUI is to serve as a screen for feigning, several important inferences can be made in regards to its use. First, participants in the present study who were

asked to feign produced lower FUI scores than honest responding participants on both the subtle and obvious items, suggesting that there are differences in performance based on response style. Additionally, although the FUI scores were lower for feign participants, on average the scores did not fall below the 50% mark that would be suggested by symptom validity testing. Consistent with previous research on forced-choice tests, this cutoff score is not sensitive enough for accurate detection. Some feigning participants produced scores that were similar to that of honest responders, suggesting that the floor effect cutoff score (to be determined in future studies) would need to be much higher than the 50% mark for an accurate detection rate. The floor effect value is necessary to evaluate the sensitivity, specificity, positive and negative predictive power of the FUI, all of which are necessary statistics before the FUI could be considered a valid screen for feigning.

Second, completion time was only partially supported as a measure of feigning for the FUI. Contrary to what was hypothesized, the majority of participants took less time to complete the FUI when feigning. There has only been one study which included completion time as a measure of feigning in the assessment of competence related knowledge and their findings are in contrast with the current research. This may be because the participants in the previous study were forensic samples, whereas the current study utilized a simulation study with college students. As college students are adept at test taking, their performances may differ from that of other populations.

The current findings on completion time do not necessarily negate the use of completion time as a screen for feigning, but instead place a different perspective on its interpretation. For example, if the results of the current research were replicated in other studies and with other populations, re-evaluating the theory of completion time may be necessary. As previously

mentioned, a hypothesis for the explanation of these results is that participants who are feigning may attempt to complete the instrument in a quick and haphazard manner and thus completion time is less than it would be in an honest responder who attends to the content of each item. This hypothesis is speculative, as no empirical data exists to support this notion, and these questions were not explored in the current research. Evaluating the use of FUI completion time in a forensic sample could provide further information into the use of this feigning detection method.

In sum, there is preliminary evidence to support the reliability and validity of the FUI within a student sample as a measure of competence related knowledge and as a potential screen for feigning. There is also support for continued research on the utilization of varying item difficulty within the FUI. Even with the previous findings, limitations exist on the interpretation of the data. It is hypothesized that future research studies will remediate some of these restrictions.

Limitations

There are several significant limitations to the current research. First, the sample utilized in Study 5 was from an undergraduate college where the focus of study is on careers in criminal justice. As such, this sample likely entered the study with a greater amount of legal related knowledge than the average person in the community would. Had the sample been obtained from an undergraduate institution without a criminal justice focus, perhaps the overall scores from honest responders would have been lower, and the items would have had higher difficulty values. An additional limitation of the sample is that, when broken down by response condition, sample size was not as large as recommended for some of the analyses conducted (i.e., factor analysis). Larger samples are necessary in future studies. In addition to the sample characteristics and size limitations, experimental sessions were run in groups of 5-10 persons. It is possible that this

influenced participant's completion time, in that observing other participants completing the instruments at earlier times may have prompted the remainder of the participants to complete the instrument quickly. Had the experimental sessions been run on an individual basis, the results may have differed.

Second, the external validity and generalizability of the results obtained in Study 5 are quite limited. Levels of motivation (e.g., feigning to obtain hospital time instead of jail time) are significantly higher, as are the stakes for feigning in a believable manner, in forensic settings. This motivation was difficult to implement in a simulation study and is an inherent flaw in this type of research design. A potential remedy to increase motivation as implemented by other researchers has been to offer monetary incentives to participants who avoid feigning detection. This method was not used in Study 5; however, participants largely endorsed putting forth effort into their performance on the instruments suggesting that some motivation was present. Additional factors that limit the external validity of the results are the influence of situational stress related to the interaction with the legal system that actual defendants are under and the potential influence of mental illness on one's performance on a measure of legal knowledge. These factors could not be assessed in a college student sample. It is not yet known how the FUI would perform in forensic contexts; however as the primary focus of this research was the conceptualization and construction of a new instrument, a simulation study was chosen because of the reduced cost and also to avoid the difficulty of implementing the measure with actual defendants at such a preliminary stage.

An additional limitation to the current research is that the instrument used to estimate intelligence in Study 5 was a screen and not a comprehensive assessment of intelligence. Intelligence levels may have been more closely approximated if full assessments had been

implemented. Current and more accurate measures of intelligence that provide an encompassing picture of intellectual functioning are needed (e.g., WAIS-IV; Wechsler, 2008). In addition to the limitations of the instrument, the sample in Study 5 consisted of participants without significant cognitive deficiencies. It is unknown how honest and feigning responders with borderline intellectual functioning would perform on the FUI. Furthermore, as the sample was college students, the majority of participants fell within the average range of intellectual functioning and only a small percentage of participants had below average intelligence, limiting the interpretation of the results from participants with lower intelligence levels.

It is possible that in persons with borderline intellectual functioning, difficulties would arise in comprehending and responding to the more challenging FUI items (ex. FUI2). However, the reading level of the FUI is comparable to the CAST-MR, an instrument designed for use with defendants who have been diagnosed with mental retardation. Thus, it is speculated that defendants of lower intelligence levels would be able to understand the FUI items. An additional concern surrounding the use of the FUI with persons with low intellectual ability is that the instrument may have to be adapted to an oral presentation of the items, if respondents are illiterate (as is typically done with the CAST-MR). If this were the case, the psychometric properties would have to be investigated when utilizing this administration format.

A final limitation of the FUI is that the instrument was not designed to assess a defendant's ability to consult with counsel or their rational understanding of the legal proceedings. As previously discussed, the rational understanding prong of *Dusky v. United States* (1960) is difficult to capture in a standardized instrument, and typically evaluators must obtain this information from the clinical interview. From the information obtained during the literature review and critical incidents phase of test development, one of the most important factors

identified in assessing factual knowledge of the legal system was the defendant's knowledge of their specific charges. The FUI does not tap into this information, as it was designed to be administered, scored, and interpreted on a nomothetic rather than individual level. These factors would have to be assessed by other means. However, it is of importance to note that information on the defendant's knowledge of their charges can be obtained from an interview, whereas the assessment of feigning is often more difficult to make solely from a clinical interview.

Future Directions

Further studies with a known-groups sample are necessary (i.e., with defendants who are genuinely lacking legal related knowledge). Scores on the FUI of incompetent defendants who genuinely lack knowledge of the legal system are needed. These responses should then be compared to the responses of a group of suspected feigners as determined by pre-administration of another feigning assessment instrument such as the SIMS (Smith & Burger, 1997). This group comparison is important, as the FUI should be sensitive to detecting feigned incompetence, but not genuine incompetence.

As the FUI was designed to detect a type of feigning that has not typically been investigated with previous instruments, difficulty may arise in obtaining a valid sample. In a genuinely incompetent sample, it cannot be assumed that defendants were found incompetent due to a lack of factual knowledge; rather they may have lacked in one of the other two areas of competency to stand trial (i.e., rational understanding or ability to consult with counsel). It would have to be ensured that the defendants in this known-groups study indeed lack legal related knowledge before the FUI could be evaluated. Even though this is a potential research obstacle, a known-groups sample would greatly improve the external validity of the FUI and help to establish potential cutoff values for the assessment of feigning.

An additional goal in the future research on the FUI is to refine the instrument to have fewer items. An initial goal in the conceptualization and construction of the FUI was to create a relatively brief instrument that could be administered quickly to a defendant. As this research was the first step in construction the instrument, the FUI is longer than would be ideal for an actual evaluation. With the refinement of the instrument will come additional psychometric analyses to determine how the reduction of the number of items impacts the reliability and validity. As previously discussed, before discarding items of similar context (e.g., several items ask about the job of the defense attorney), it would be of interest to investigate consistency in item endorsement.

Finally, it is of importance to continue investigating the relationship between intelligence level and item type in participants with different response styles. As the current study showed preliminary support for the hypothesized relationship between the three variables, participants with wider ranges of intellectual functioning need to be investigated if these variables are to be used as a mechanism of differentiating between responders.

Conclusion

The preliminary construction and pilot testing of an instrument designed to assess factual understanding of the legal system while simultaneously screening for feigning produced some promising results. With further research into the use of the FUI in forensic settings the instrument could add to the resources mental health professionals have when assessing criminal defendants. As the FUI items are applicable across the majority of jurisdictions within the United States, the instrument could be useful in many contexts.

It is important to note that the FUI was designed to be a screen for feigning rather than a comprehensive assessment tool. When feigning screens such as the FUI raise questions about a

defendant's response style, further evaluation is necessary before a determination into the their presentation as feigning can be made. Likewise, if the screens do not raise questions about a defendant's response style, one can be more confident that the defendant is answering honestly and that the results of the examination reflect his or her true abilities.

Even with useful screening instruments, mental health professionals must always incorporate clinical judgment and a clinical interview, and a thorough review of records into their assessments. It is pertinent that the professional ensure that the defendants portrayal is not due to genuine psychopathology or genuine cognitive deficiency (e.g., psychoses or mental retardation), but rather can be attributed to intentional distortion of their abilities. An instrument should never be used in isolation to determine whether or not a person is feigning. Because of the grave consequences a defendant faces when incorrectly labeled as feigning (i.e., denial of psychoeducation or other mental health services), diagnostic accuracy is necessary when making such determinations. Future explorations into the accuracy of the FUI in the detection of feigning and in accurately assessing factual knowledge of the legal system will help to improve upon the instruments currently available to mental health professionals.

Appendix A

Factual Understanding Instrument (FUI) Items Created During Study 3

1) What are some legal rights defendants have when they are arrested and questioned by the police?
    a. Right to remain silent, right to have an attorney
    b. Right to speak to the police, right to take back statements made to the police

2) What are some of the legal rights defendants have when they go to court?
    a. Defendants have no rights when they go to court because they have to do what the authorities tell them to do
    b. Defendants have the right against self-incrimination, the right to be represented by a lawyer, the right to confront witnesses

3) What does it mean for a person to waive their rights?
    a. The person gives up those rights
    b. The person waives their hand at someone

4) What happens at an arraignment?
    a. The person is formally charged with a crime and enters a plea
    b. The person argues their case with the help of their attorney

5) What does it mean to be charged with a crime?
    a. To be accused of an illegal act
    b. To pay a fine or serve jail time

6) What is a crime?
    a. An attempt to earn extra money
    b. An act in which a person breaks the law

7) What causes a person to be charged with a crime?

    a. The person is accused of doing something illegal

    b. The person makes a phone call

8) What does it mean to be arrested?

    a. The police take a person into custody

    b. The judge takes a person to rest

9) If a person talks to the police when arrested, what do police do with that information?

    a. The police will not report the information because they are there to help

    b. The police will report the information and it may be used in court

10) Why is the police's version of the crime important?

    a. Because their version may be used in court to prove the defendant's guilt

    b. Because their version may be used to help the defendant get a job

11) What does it mean to have the right to remain silent when arrested?

    a. The person has the right to not answer basic questions when they are arrested (e.g., name, date of birth, address)

    b. The person has the right to not answer any questions beyond basic information when arrested (e.g., questions about the crime)

12) What does it mean to have the right to an attorney?

    a. Defendants have the option to have an attorney help them during police questioning and during the trial

    b. Defendants have the option to have an attorney help them commit further crimes

13) What is a defense attorney?

    a. An attorney who represents the judge

    b. An attorney who represents the defendant

14) What does the attorney for the defense do?

   a. Helps the defendant win the case
   b. Harms the people of the jury

15) What is the job of the defense attorney?

   a. To help the defendant
   b. To help the government

16) Whose side does the defense attorney work for?

   a. The prosecution
   b. The defendant

17) Who is the prosecutor (also called the district attorney)?

   a. The attorney who tries to prove the defendant is guilty of the crime
   b. The person who sits in the courtroom and types what is being said

18) What is the job of the prosecutor/district attorney?

   a. To prove the defendant is a nice person
   b. To prove the defendant committed the crime

19) What does the prosecutor or district attorney do?

   a. Tries to prove the defendant is guilty of the crime
   b. Tries to protect the judge from the defendant

20) Whose side is the prosecutor on?

   a. The side that is against the defendant
   b. The side that earns the most money

21) Who is the judge?

a. The person who is in charge of the courtroom

b. The person who calls on witnesses to testify

22) What is a job of the judge?

a. To make sure the trial is fair

b. To report on the weather

23) Who is in charge of the court?

a. The judge

b. The police

24) What is a job of the jury?

a. To make sure each side gets to argue their case in court

b. To make a decision about the guilt or innocence of the defendant

25) What is a job of the jury?

a. To instruct the judge on truthfulness of the evidence

b. To listen to both sides of the case and to make a fair decision

26) How many jurors must believe the defendant is guilty in order for the defendant to be found guilty?

a. All of them

b. Most of them

27) How convinced must a jury be in order to say a person is guilty of a crime?

a. They should be at least 50% convinced

b. They should be convinced beyond a reasonable doubt

28) What does “beyond a reasonable doubt” refer to?

a. It is the threshold which must be reached before a defendant can be found guilty

b. The way a defendant feels when he/she is unsure of the outcome of the trial

29) From where are the jurors used in court cases selected?

a. They are selected from the prison inmates

b. They are selected from the general community

30) How many people are on a trial jury?

a. Twelve

b. Fifteen

31) What is a hung jury?

a. When a person decides that he/she no longer wants to be on the jury

b. When the jury cannot reach a decision about the defendant’s guilt or innocence

32) What happens after a hung jury occurs?

a. The defendant will go free

b. The case will be retried

33) What is the job of the defendant?

a. To work with his/her attorney to resolve the legal case

b. To stay out of further legal trouble until the case is resolved

34) What is the job of a witness?

a. To report to the court what they saw during the crime

b. To spend time in jail for committing a crime

35) Who says in court that they have firsthand knowledge of the defendant's relationship to the crime?

    a. The prosecutor

    b. The witness

36) What does it mean to plead?

    a. The defendant makes a statement to the court saying whether they are guilty or innocent of the crime

    b. The defendant makes a statement to the court asking for more time to prepare the case

37) What does it mean to plead guilty?

    a. The defendant says he/she committed the crime

    b. The defendant tells the court that he/she lives alone

38) What does it mean to plead not guilty?

    a. The defendant says that he/she did not commit the crime

    b. The defendant says he/she has no criminal record

39) What does it mean when a defendant says he/she is innocent?

    a. The defendant says he/she has good morals and values

    b. The defendant says that he/she did not commit the crime

40) What does it mean to plead no contest?

    a. The defendant says that he/she does not want to argue the charges in court

    b. The defendant says that he/she wants to serve time in prison

41) What happens after a defendant pleads guilty?

    a. The defendant will argue their case in court and the jury will decide on guilt or innocence

b. The defendant will not be able to argue their case in court and he/she will be punished for the crime

42) What happens after a defendant pleads not guilty?

a. There will be a trial and both sides will argue for the defendant's innocence

b. There will be a trial in which one side argues for the defendant's guilt and the other side argues for innocence

43) If a defendant pleads guilty he/she will give up what legal right?

a. The right to argue the evidence put forth by the prosecution

b. The right to be represented by a defense attorney

44) What does it mean to plead not guilty by reason of insanity?

a. The defendant did not commit the crime and therefore is not guilty

b. The defendant was incapable of knowing what he/she did was wrong when they committed the crime

45) Who would a defendant want to speak to before pleading guilty?

a. His/her own attorney

b. The local media

46) What does it mean to be convicted?

a. The court finds that the defendant is guilty of the crime

b. The defendant tells the court he/she is guilty of the crime

47) What does it mean to be acquitted?

a. To be found not guilty of the crime

b. To quit talking to the attorney

48) Where does a person go if they are found not guilty by reason of insanity?

a. Back to the community

b. A psychiatric hospital

49) What does “time served” mean?

a. The time the defendant has spent in jail for prior crimes

b. The time the defendant has spent in jail before the court decided on guilt

50) What is a sentence?

a. The time it takes the defendant to get to the courthouse

b. The punishment the defendant receives for his/her crime

51) What are common punishments for someone who is found guilty of a crime?

a. Jail, probation, fines, community service

b. School, counseling, exercise

52) What is a maximum sentence?

a. The largest punishment a defendant may get if found guilty

b. The most likely punishment that a defendant may get if found guilty

53) What is a minimum sentence?

a. The least amount of knowledge a person must have to graduate

b. The smallest punishment a defendant may get if found guilty

54) What is a verdict?

a. The decision reached by the jury about the defendant’s guilt or innocence

b. The time used by the jury to discuss and consider all the evidence of the trial

55) What is jail/prison?

a. A place where people go to as a means of punishment for committing a crime

b. A place where people go for vacation

56) What is probation?

a. When a person is locked in a single cell for a lengthy period of time

b. When a person in placed under supervision and is closely monitored while living in the community

57) What is a fine?

a. A sum of money the defendant must pay as punishment for the crime

b. A sum of money the defendant owes to his/her attorney

58) Which punishment is more severe?

a. Six month in jail

b. Six hours of community service

59) In court, who is working to prove the defendant is guilty?

a. The teacher

b. The prosecutor

60) In court, who will try to argue that the defendant is not guilty?

a. The defense attorney

b. The corrections officer

61) Who is able to argue against what the prosecution says in court?

a. The judge

b. The defense attorney

62) After the prosecutor and defense attorney present the evidence, what happens next?

a. The jury meets and they reach a decision on the guilt or innocence of the defendant

b. The prosecutor and defense attorney discuss the evidence amongst themselves

63) After the jury decides a defendant is guilty, what happens next?

a. The defendant is set free and has no punishment

b. The judge or jury decides on the defendant's punishment

64) After the jury decides a defendant is not guilty, what happens next?

a. The defendant is cleared of the charges and is not punished

b. The defendant receives a serious punishment

65) What does it mean to be innocent until proven guilty?

a. The defendant has to be proven innocent in court

b. The defendant has to be proven guilty in court

66) When the court gives the defendant a punishment, does the defendant have to follow the punishment?

a. Yes, the defendant always has to follow the punishment because the court has power over the defendant

b. No, the defendant never has to follow the punishment because the court is trying to scare the defendant

67) What should defendants do while they are in court?

a. Speak to the prosecuting attorney about the case

b. Follow the directions of their defense attorney

68) Why might the prosecutor ask the defendant questions while in court?

a. The prosecutor is trying to get information to reduce the defendant's punishment

b. The prosecutor is trying to get information to make the defendant look guilty

69) What does it mean to testify?

a. To make a statement in front of the court for the purposes of establishing the truth

b. To make a statement in front of the court to make the defendant appear guilty

70) If a defendant testifies, who will ask him/her questions?

a. The audience, the court reporter, and possibly his/her family

b. The defense attorney, the prosecutor, and possibly the judge

71) When a defendant testifies, what will he/she have to say?

a. He/she will have to say that they committed the crime

b. He/she will have to answer questions asked by the attorneys

72) What is a plea bargain?

a. An agreement in which the defendant pleads not guilty to additional charges after a trial

b. A settlement of the case where the defendant pleads guilty to a lesser charge

73) What is the defendant saying he/she did when they take a plea bargain?

a. The defendant is saying he/she is guilty of a lesser charge than what originally charged with

b. The defendant is saying he/she is guilty of only one misdemeanor charge

74) What would happen if a defendant took a plea bargain and later he/she was unhappy about it?

a. The defendant would not have to complete the sentence because there would be a new trial

b. The defendant would not be able to argue the case again at a later date because they already took the plea

75) What happens if a defendant does not take a plea bargain that has been offered?

a. The defendant's case will go to court where a jury will decide on his/her guilt or innocence

b. The defendant will lose their right to a jury trial and their right to an attorney

76) What is a misdemeanor?

a. A minor crime that carries a punishment of up to one year jail time or a fine

b. A minor crime that carries a punishment of up to five years jail time or a fine

77) What is a felony?

a. A person who saves the life of a sick or injured animal

b. A serious crime that can be punished by a lengthy prison term

78) What is evidence?

a. The time when the defendant has a visit from family members

b. Something used to demonstrate the truth during a trial

79) What can be used as evidence in court?

a. Information that is related to the defendant or the legal case

b. Any information from any person who lives near the courthouse

80) How can evidence be used in court?

a. To support the prosecutor's side and not the defendant's side

b. To support either the prosecutor's side or the defendant's side

81) How is evidence used *against* the defendant in court?

a. It can be used to help support the prosecutor's case

b. It can be used as a reward for the defendant's good behavior

82) What does it mean to be incompetent to stand trial?

a. The defendant is unable to understand the legal process and will not be able to immediately move forward with his/her case

b. The defendant is unable to understand the legal process has and therefore will be set free and cleared of the charges

83) What is a hearing?

a. A court procedure where the initial evidence and charges against the defendant are reviewed

b. The ability to listen very well to someone who is talking

84) What is a trial?

a. The process of deciding on a strategy with one's attorney before going to court

b. The process of going to court in order to determine a defendant's guilt or innocence

85) What is an oath?

a. A promise made in front of the court where the person says they will tell the truth

b. A promise made by the defendant to his/her attorney that they will tell the truth

86) What is a bench trial?

a. A trial in which there is no jury and the judge decides the guilt or innocence of the defendant

b. A trial in which there are two separate juries who decide on the defendant's guilt or innocence

87) What is a defense strategy?

a. The plan the defense attorney and the defendant decide to use to help win the case

b. The plan the defendant has to commit crimes when he/she is released from jail

88) What is an appeal?

a. When the defendant's case is retried by a higher court

b. When the defendant's case is placed on hold

Appendix B

AP-LS Listserve Recruitment Announcement

Dear Mental Health Professional,

I am a doctoral student in the Department of Psychology at John Jay College of Criminal Justice currently working on my dissertation. I am recruiting mental health professionals who routinely conduct competency to stand trial evaluations or scientific researchers who specialize in this area to participate in an expert review process.

The aim of the dissertation research is to construct an instrument aimed at capturing particular aspects of a defendant's competence-related abilities and a crucial part of the instrument's construction is to have experts in the field rate the quality of the items before they are tested in a pilot study. Eligible persons for this study are those who have experience in assessing criminal defendants for their competence-related abilities or those who have researched and/or published articles on the topic.

Participation in this study will involve reading approximately 40 items and rating each one on a list of characteristics. The study can be completed by hand (materials will be mailed through postal mail) or can be e-mailed for a faster delivery and completed electronically. The review process will require approximately 1 hour of your time.

If you are interested in participating in this study or would like to obtain further information, please contact Elizabeth Arias at earias@gc.cuny.edu or dissertation supervisor Dr. Michele Galietta at mgalietta@jjay.cuny.edu. This study has been approved by the John Jay College of Criminal Justice Institutional Review Board (IRB).

Thank you for your time and any help you can offer.

Sincerely,

Elizabeth Arias

Appendix C

Instructions and Review Form Example for Expert Reviewers

Dear Dr. X,

Thank you for agreeing to participate in the expert review of items for a new prospective instrument aimed at assessing a defendant's factual knowledge of the legal system. Attached you will find a consent form and a demographics form. Please fill out each and return with the review forms.

Instructions for review process: Attached you will find 44 items to be reviewed. The following represent a sample of the larger pool of items that were constructed for the study. Each item is printed on its own page. Each page will have an item printed at the top consisting of a question and two answer choices. One answer is correct and one is incorrect. The items purposely vary in their difficulty level from very easy to difficult. Below the item are six item characteristics for you to judge, each on a 5-point Likert scale. If you rate an item lower than a 3 on any characteristic please provide as much information as possible as to what you believe is problematic about the item. A space beneath the Likert scales is designated for this information. You may also use this space to make general comments about the item.

After you complete the review of the items, please see the attached debriefing form and return the materials to me within 2 weeks or at your earliest convenience.

Thank you for your time and participation.

Sincerely,

Elizabeth Arias

Item #1:

Question: What does it mean to be arrested?

Answer 1: The police take a person into custody

Answer 2: The police give a person a parking ticket

For this item, please rate the:

a) difficulty of the question for those with a laypersons understanding of the legal system	☐ very easy	☐ easy	☐ average	☐ difficult	☐ very difficult
b) grammar/wording of question	☐ very poor	☐ poor	☐ acceptable	☐ good	☐ very good
c) grammar/wording of responses	☐ very poor	☐ poor	☐ acceptable	☐ good	☐ very good
d) accuracy of correct response	☐ very poor	☐ poor	☐ acceptable	☐ good	☐ very good
e) plausibility of incorrect response	☐ very poor	☐ poor	☐ acceptable	☐ good	☐ very good
f) overall quality of the item	☐ very poor	☐ poor	☐ acceptable	☐ good	☐ very good

For this item, please provide information for any rating given below average/acceptable and any additional comments you have. Please be as specific as possible:

_____
_____
_____
_____
_____

Appendix D

Factual Understanding Instrument (FUI)

Instructions: Please circle one answer for each item

1) What does it mean to be arrested?
    a. The police take a person into custody
    b. The police give a person a parking ticket

2) What happens at an arraignment?
    a. The person is formally charged with a crime and may enter a plea
    b. The person argues his/her case with the help of an attorney

3) What are some of the legal rights defendants have when they go to court?
    a. Defendants have no rights when they go to court because they have to do what the authorities tell them to do
    b. Defendants have the right to be represented by an attorney, the right to face witnesses, the right to not testify

4) What is the job of the prosecutor/district attorney?
    a. To prove the defendant committed the crime
    b. To determine the guilt or innocence of the defendant

5) Who is the judge?
    a. The person who is in charge of the courtroom
    b. The person who charges the defendant with a crime

6) What does it mean when a defendant says he/she is not guilty?
    a. The defendant says he/she has good morals and values
    b. The defendant says that he/she did not commit the crime

7) What is evidence?

   a. The time when the defendant has a visit from family members

   b. Something used to demonstrate the truth during a trial

8) What does it mean to have the right to an attorney?

   a. Defendants have the option to have an attorney help them during police questioning and during the trial

   b. Defendants have the option to have an attorney help them commit further crimes

9) What is a job of the jury?

   a. To instruct the judge on truthfulness of the evidence

   b. To listen to both sides of the case and to make a fair decision

10) What is a crime?

   a. A place where the judge and attorneys meet

   b. An act in which a person breaks the law

11) How many jurors must believe the defendant is guilty in order for the defendant to be found guilty?

   a. All of them

   b. Some of them

12) What does it mean for a person to waive his/her rights?

   a. The person gives up those rights

   b. The person waives his/her hand at someone

13) What does it mean to be charged with a crime?

   a. To be accused of an illegal act

   b. To pay a fine or serve jail time

14) What does “beyond a reasonable doubt” refer to?

a. The level of certainty that must be reached before a defendant can be found guilty

b. The way a defendant feels when he/she is unsure of the outcome of the trial

15) What does the prosecutor or district attorney do?

a. Tries to prove the defendant is guilty of the crime

b. Tries to protect the judge from the defendant

16) What does it mean to be acquitted?

a. To be found not guilty of the crime

b. When the attorney withdraws from the case

17) What is a hung jury?

a. When a person decides that he/she no longer wants to be on the jury

b. When the jury cannot decide whether or not the defendant is guilty

18) What is the most important job of the defendant?

a. To work with his/her attorney to resolve the legal case

b. To speak to other inmates about his/her legal case

19) What is the job of a fact witness?

a. To report to the court what he/she saw or heard during the crime

b. To decide whether the defendant is guilty or not guilty

20) What does it mean to plead to a charge?

a. The defendant tells the court whether he/she is guilty or not guilty of the charge

b. The defendant asks the court for more time to prepare the case

21) What is meant by the term “maximum sentence?”

a. The largest punishment a defendant may get if found guilty

b. The most likely punishment that a defendant may get if found guilty

22) What is the job of the defense attorney?

a. To help the defendant

b. To help the government

23) Which punishment is more serious?

a. Six months in jail

b. One month of probation

24) How can evidence be used in court?

a. To support the prosecutor’s side and not the defendant’s side

b. To support either the prosecutor’s side or the defendant’s side

25) What happens after a defendant pleads guilty?

a. The defendant will argue the case in court and the jury will decide on guilt or lack of guilt

b. The defendant will not be able to argue the case in court, and he/she will be punished for the crime

26) Why is the police version of the crime important?

a. Because the police version may be used in court to prove the defendant’s guilt

b. Because the police version may be used to help the defendant get a job

27) What does it mean to plead not guilty by reason of insanity?

a. The defendant did not commit the crime and therefore is not guilty

b. The defendant was incapable of knowing what he/she did was wrong when he/she committed the crime due to a mental illness.

28) What does it mean to be convicted?

a. The court finds that the defendant is guilty of the crime

b. The court finds the defendant not guilty of the crime

29) What is a defense attorney?

a. An attorney who represents the judge

b. An attorney who represents the defendant

30) Who is more important for a defendant to speak with before pleading guilty?

a. His/her own attorney

b. The local TV station

31) What does “time served” mean?

a. A sentence that requires a person to attend counseling and be monitored while living in the community

b. A sentence that is equal to the time the defendant spent in jail while waiting for the court to decide on guilt

32) Who testifies about personal knowledge related to the crime (what he/she saw or heard directly)?

a. The prosecutor

b. The fact witness

33) Where are jury members selected from?

a. The prison inmates

b. The general community

34) If a defendant pleads guilty, he/she will give up what legal right?

a. The right to argue against the prosecutor’s case

b. The right to be represented by a defense attorney

35) What is a sentence?

a. The time it takes the defendant to get to the courthouse

b. The punishment the defendant receives if found guilty

36) Typically, how many people are on a jury in a criminal trial?

a. Twelve

b. Fifteen

37) If a defendant testifies, who will ask him/her questions?

a. The audience, the court reporter, and possibly his/her family

b. The defense attorney, the prosecutor, and possibly the judge

38) What is a verdict?

a. The decision reached by the judge or jury about the defendant's guilt

b. The time used by the judge or jury to discuss and consider all the evidence of the trial

39) What is a jail/prison?

a. A place where people go as a means of punishment

b. A place where people go for vacation

40) What is meant by the term "minimum sentence?"

a. The largest punishment a defendant may get if found guilty

b. The smallest punishment a defendant may get if found guilty

41) What does it mean to plead not guilty?

a. The defendant says that he/she did not commit the crime

b. The defendant says he/she has no criminal record

42) What is a hearing?

a. A preliminary court appearance to decide on some matter related to the case

b. The ability to listen very well to someone who is talking

43) Who is in charge of the court?

a. The judge

b. The police

44) What is probation?

a. When a person is locked in a single cell for a lengthy period of time

b. When a person in placed under supervision and is monitored closely while living in the community

45) What is a fine?

a. A sum of money the defendant must pay as punishment for the crime

b. A sum of money the defendant owes to his/her attorney

46) In court, who is trying to prove that the defendant is guilty?

a. The judge

b. The prosecutor

47) What is a job of the jury?

a. To make sure each side gets to argue its case in court

b. To make a decision about whether or not the defendant is guilty

48) What happens after all of the evidence in a criminal case is presented?

a. The jury meets and they reach a decision on the guilt or innocence of the defendant

b. The prosecutor and defense attorney discuss the evidence amongst themselves

49) What does it mean to plead guilty?

a. The defendant says he/she committed the crime

b. The defendant says that someone else committed the crime

50) Who will argue against what the prosecutor says in court?

a. The police officer

b. The defense attorney

51) After the jury decides a defendant is guilty, what happens next?

a. The defendant is set free and has no punishment

b. The judge decides on the defendant's punishment

52) What is one job of the judge?

a. To make sure the trial is fair

b. To report on the weather

53) What does it mean to be innocent until proven guilty?

a. The defendant has to be proven innocent in court

b. The defendant has to be proven guilty in court

54) What should defendants do while they are in court?

a. Follow the directions of their defense attorney

b. Speak to the witnesses about the case

55) Why might the prosecutor ask the defendant questions while in court?

a. The prosecutor is trying to get information to reduce the defendant's punishment

b. The prosecutor is trying to get information to make the defendant look guilty

56) Which side is the prosecutor on?

a. The defense

b. The state/government

57) When a defendant testifies, what will he/she have to say?

a. He/she will have to say that he/she committed the crime

b. He/she will have to answer questions asked by the attorneys

58) What is a plea bargain?

a. The defendant pleads not guilty to a charge and the case goes to trial

b. A settlement of the case where the defendant pleads guilty to a charge

59) What is a misdemeanor?

a. A crime that carries a punishment of up to one year jail time or a fine

b. A crime that carries a punishment of up to five years jail time

60) In court, who will argue that the defendant is not guilty?

a. The defense attorney

b. The judge

61) Where do most people go immediately after being found not guilty by reason of insanity?

a. Back to the community

b. A state psychiatric hospital

62) What is a felony?

a. A minor violation such as a speeding ticket that is punished with a fine

b. A serious crime that can be punished by a prison term of more than one year

63) What is an appeal?

a. When the defendant's case is reviewed by a higher court

b. When the defendant's case is placed on hold

64) What can be used as evidence in court?

a. Information that is related to the defendant or the legal case

b. Any information from any person who lives near the courthouse

65) What happens after a defendant pleads not guilty?

a. There will be a trial, and both sides will argue the defendant is not guilty

b. There will be a trial in which one side argues the defendant is guilty, and the other side will argue the defendant is not guilty

66) How convinced must a jury be in order to say a person is guilty of a crime?

a. Members should be at least 50% convinced

b. Members should be convinced beyond a reasonable doubt

67) Which side does the defense attorney work for?

a. The prosecution

b. The defendant

68) What is the defendant saying when he/she takes a plea bargain?

a. The defendant is admitting to committing a crime

b. The defendant is requesting a new attorney

69) What is a trial?

a. The process of deciding on a strategy with one's attorney before going to court

b. A judicial process of going to court to determine a defendant's guilt

70) After the jury decides a defendant is not guilty, what happens next?

a. The defendant is cleared of the charges and is not punished

b. The defendant receives a serious punishment

71) What is an oath?

a. A promise made in front of the court where the person says he/she will tell the truth

b. A promise made by the defendant to his/her attorney that he/she will tell the truth

72) How is evidence used against the defendant in court?

a. It can be used to prove the defendant is guilty

b. It can be used as a reward for the defendant's good behavior

73) What would happen if a defendant accepted a plea bargain and later he/she was unhappy about it?

a. He/she would not have to complete the punishment and there would be a new trial

b. The defendant would not be able to change his/her mind, because he/she already pled guilty

Appendix E

Instructions for Feign Participants

For the next part of the study, I would like each of you to pretend that you have been charged with a crime and that you will be going to trial.

But before you go to trial you will be given two tests that are often given to people who are going to court for a criminal trial to see what they know about the legal system.

On these tests I want you to answer the questions as if you have a lack of knowledge of the legal system and how it works. You are to pretend as though you do not know the basic "players" of the courtroom or basic courtroom procedures. Your job is to fool the test, by pretending that you do not have this knowledge, so that you do not have to go to trial or face a prison sentence.

However, just answering all the questions wrong on the tests won't work because most people know at least some of these answers. This approach is too obvious and if you are caught trying to fake being impaired your prison sentence may be even longer. Therefore, it is important that you fake in a way that is believable, but not too obvious.

One test that you will take has 73 questions and each will have two answers – one right and one wrong. The other test will have 40 questions with three possible answer choices. Only one of them is correct.

You will need to answer all the questions on these two tests and your goal is to look impaired to the point that you do not know the basic roles or the "players" in the legal process and concepts involved in the courtroom, but not so impaired that it is obvious that you are faking. Please be consistent in your responding across both of the tests.

On each test there will be a space at the top for you to record the time you started the test. Please look at the clock and write the time you started the test when you begin. Please do this for both of the tests. At the end of each test there will also be a space for you to record the time you finished the test. Please write the time in when you finish each test. Although your time will be recorded, there is no time limit for these tests.

Also, when you complete each of the two tests, please raise your hand so the experimenter can also record the time you finished the test. When you finish the first test please sit quietly until the rest of the group has finished their test before starting the second test. Everyone has to start the second test at the same time. Please wait for instructions from the experimenter before beginning the second test. There will be reminders for all of this information in your packets.

Do you understand these instructions? Are there any questions?

Can I provide any additional information to help you follow these instructions before you begin?

Instructions for Honest Participants

For the next part of the study, I would like each of you to pretend that you have been charged with a crime and that you will be going to trial.

Before you go to trial you will be given two tests that are often given to people who are going to court for a criminal trial to see what they know about the legal system.

On these tests I want you to answer the questions to the best of your ability.

One test will have 73 questions and each will have two answers – one right and one wrong. The other test will have 40 questions with three possible answer choices. Only one of them is correct. You will need to answer all the questions on both tests and your goal is to answer the items correctly.

On each test there will be a space at the top for you to record the time you started the test. Please look at the clock and write the time you started the test when you begin. Please do this for both of the tests. At the end of each test there will also be a space for you to record the time you finished the test. Please write the time in when you finish each test. Although your time will be recorded, there is no time limit for these tests.

Also, when you complete each of the two tests, please raise your hand so the experimenter can also record the time you finished the test. When you finish the first test please sit quietly until the rest of the group has finished their test before starting the second test. Everyone has to start the second test at the same time. Please wait for instructions from the experimenter before beginning the second test. There will be reminders for all of this information in your packets.

Do you understand these instructions? Are there any questions?

Can I provide any additional information to help you follow these instructions before you begin?

Appendix F

Table F1

FUI Means, Standard Deviations, and Point-Biserial Correlations

	Combined[a]			Honest[b]			Feign[c]		
Item	*M*	*SD*	r_{pb}	*M*	*SD*	r_{pb}	*M*	*SD*	r_{pb}
FUI1	.96	.19	.37	1	0		.92	.27	.37
FUI2	.45	.50	.35	.67	.47	.29	.25	.43	-.04
FUI3	.71	.46	.70	.99	.10	.49	.43	.50	.46
FUI4	.59	.49	.33	.74	.44	.33	.44	.50	.14
FUI5	.76	.43	.28	.84	.37	.21	.68	.47	.24
FUI6	.84	.37	.54	.99	.10	.04	.70	.46	.43
FUI7	.88	.32	.55	1	0		.77	.42	.48
FUI8	.90	.30	.49	1	0		.81	.39	.42
FUI9	.66	.48	.66	.97	.17	.05	.36	.48	.38
FUI10	.91	.28	.52	1	0		.83	.37	.49
FUI11	.60	.49	.49	.83	.38	.02	.38	.49	.31
FUI12	.81	.40	.60	.99	.10	.08	.63	.49	.47
FUI13	.57	.50	.64	.90	.30	.12	.26	.44	.35
FUI14	.61	.49	.77	.96	.20	.31	.28	.45	.54
FUI15	.80	.40	.61	.98	.14	.49	.62	.49	.47
FUI16	.58	.49	.37	.79	.41	.16	.39	.49	.09
FUI17	.62	.49	.73	.96	.20	.19	.29	.46	.48
FUI18	.84	.37	.56	.99	.10	-.03	.70	.46	.47
FUI19	.77	.42	.67	.97	.17	.51	.58	.50	.56
FUI20	.71	.46	.56	.92	.27	.43	.51	.50	.38
FUI21	.77	.42	.57	.96	.20	.45	.59	.49	.40
FUI22	.89	.31	.43	.99	.10	-.10	.80	.40	.36
FUI23	.91	.29	.42	.98	.14	.03	.84	.37	.42
FUI24	.70	.46	.57	.96	.20	-.03	.44	.50	.30
FUI25	.66	.48	.38	.78	.42	.09	.54	.50	.37
FUI26	.89	.32	.51	1	0		.78	.41	.42
FUI27	.75	.43	.64	.98	.14	.07	.54	.50	.48
FUI28	.82	.39	.59	.97	.17	.29	.67	.47	.52
FUI29	.89	.32	.43	.99	.10	.49	.79	.41	.30
FUI30	.86	.35	.59	1	0		.73	.45	.51
FUI31	.69	.46	.59	.91	.29	.49	.48	.50	.40
FUI32	.74	.44	.58	.96	.20	.20	.54	.50	.40
FUI33	.91	.28	.46	.99	.10	-.08	.84	.37	.46
FUI34	.63	.48	.40	.85	.36	.12	.42	.50	.10

FUI35	.84	.37	.59	1	0		.69	.47	.48
FUI36	.62	.49	.49	.84	.37	.25	.42	.50	.28
FUI37	.82	.39	.66	1	0		.64	.48	.56
FUI38	.69	.46	.63	.95	.22	.30	.44	.50	.40
FUI39	.96	.20	.34	.98	.14	.56	.93	.25	.38
FUI40	.87	.34	.39	.95	.22	-.03	.79	.41	.40
FUI41	.71	.45	.72	.98	.14	.56	.46	.50	.52
FUI42	.65	.48	.77	.99	.10	.01	.32	.47	.54
FUI43	.86	.35	.58	.99	.10	-.01	.73	.45	.54
FUI44	.73	.44	.75	.98	.10	.40	.49	.50	.61
FUI45	.72	.45	.69	.98	.14	.20	.47	.50	.51
FUI46	.75	.43	.67	.98	.14	.54	.54	.50	.49
FUI47	.68	.47	.66	.94	.24	.42	.42	.50	.45
FUI48	.67	.47	.71	.95	.22	.45	.41	.49	.52
FUI49	.87	.33	.59	.99	.10	.06	.76	.43	.58
FUI50	.85	.36	.60	.99	.10	.49	.71	.46	.52
FUI51	.91	.29	.52	1	0		.82	.39	.48
FUI52	.95	.23	.40	1	0		.90	.31	.37
FUI53	.72	.45	.38	.86	.35	.15	.58	.50	.27
FUI54	.71	.46	.61	.89	.31	.35	.53	.50	.56
FUI55	.76	.43	.54	.95	.22	.37	.58	.50	.36
FUI56	.81	.40	.52	.94	.24	.36	.68	.47	.46
FUI57	.72	.45	.60	.95	.22	.20	.50	.50	.41
FUI58	.65	.48	.49	.92	.27	.39	.40	.49	.08
FUI59	.69	.46	.58	.90	.30	.28	.49	.50	.45
FUI60	.84	.37	.53	.96	.20	.32	.72	.45	.48
FUI61	.61	.49	.34	.72	.45	.32	.50	.50	.28
FUI62	.70	.46	.45	.86	.35	.05	.55	.50	.39
FUI63	.54	.50	.54	.82	.39	.07	.27	.45	.28
FUI64	.79	.41	.64	.98	.14	.10	.61	.49	.55
FUI65	.74	.44	.53	.94	.24	.00	.56	.50	.39
FUI66	.60	.49	.74	.94	.24	.53	.28	.45	.48
FUI67	.87	.34	.47	.96	.20	.52	.78	.41	.43
FUI68	.68	.47	.64	.90	.30	.44	.46	.50	.52
FUI69	.71	.46	.69	.95	.22	.48	.48	.50	.54
FUI70	.89	.31	.42	.97	.17	.28	.82	.39	.39
FUI71	.67	.47	.65	.94	.24	.13	.42	.50	.46
FUI72	.85	.36	.56	.97	.17	.70	.74	.44	.48
FUI73	.64	.48	.50	.84	.37	.37	.44	.50	.32

Note. *M*=mean; *SD*=standard deviation; r_{pb}=point biserial correlation.
[a]N=214. [b]n=101. [c]n=113

Table F2

FUI Item Difficulties and Discrimination under Item Response Theory

	Combined		Honest		Feign	
Item	Diff	Discrim	Diff	Discrim	Diff	Discrim
FUI1	-2.20	2.27			-1.78	2.05
FUI2	.44	.92	-.63	1.20	12.61	.09
FUI3	-.63	2.44	-2.05	28.69	.33	1.16
FUI4	-.37	.72	-1.04	1.11	1.11	.24
FUI5	-1.87	.62	-2.68	.64	-1.17	.68
FUI6	-1.30	2.08	-6.90	.68	-.82	1.32
FUI7	-1.44	2.59			-1.00	1.69
FUI8	-1.60	2.34			-1.10	1.90
FUI9	-.50	1.90	-4.72	.77	.75	.81
FUI10	-1.49	3.77			-1.04	2.62
FUI11	-.24	1.17	-4.12	.39	.95	.71
FUI12	-1.08	2.10	-4.58	1.07	-.45	1.36
FUI13	-.13	1.80	-5.54	.40	1.27	.97
FUI14	-.28	3.08	-2.20	1.85	.83	1.66
FUI15	-1.05	1.97	-1.94	3.91	-.45	1.19
FUI16	-.27	.84	-2.44	.55	3.45	.15
FUI17	-.29	2.22	-3.21	1.08	1.09	.99
FUI18	-1.31	2.00	9.02	-.53	-.74	1.55
FUI19	-.96	2.14	-1.78	3.75	-.31	1.69
FUI20	-.69	1.57	-1.65	2.02	.10	.87
FUI21	-1.04	1.77	-1.79	2.85	-.52	.95
FUI22	-1.67	1.70	2.66	-3.99	-1.23	1.27
FUI23	-1.93	1.64	-9.26	.42	-1.41	1.60
FUI24	-.71	1.43	221.40	-.01	.42	.60
FUI25	-.77	.75	-2.62	.49	-.08	.74
FUI26	-1.50	2.27			-1.02	1.65
FUI27	-.87	2.04	-3.56	1.21	-.14	1.17
FUI28	-1.16	2.03	-2.81	1.44	-.57	1.69
FUI29	-1.73	1.65	-2.05	28.69	-1.61	.92
FUI30	-1.36	2.13			-.79	1.62
FUI31	-.65	1.60	-1.25	3.59	.10	1.08
FUI32	-.90	1.58	-3.28	1.05	-.15	.95
FUI33	-1.72	1.94	3.22	-2.10	-1.12	2.02
FUI34	-.51	.90	-2.39	.76	1.98	.19
FUI35	-1.23	2.54			-.68	1.63

FUI36	-.40	1.13	-1.94	.92	.74	.59
FUI37	-1.06	2.73			-.41	1.93
FUI38	-.56	1.88	-2.19	1.67	.43	.96
FUI39	-2.32	1.99	-1.73	50.87	-1.77	2.64
FUI40	-2.18	1.03	26.80	-.11	-1.38	1.28
FUI41	-.64	2.80	-1.73	50.87	.19	1.68
FUI42	-.39	3.07	-5.17	.94	.72	1.64
FUI43	-1.36	2.28	-9.79	.47	-.80	1.94
FUI44	-.70	3.19	-2.66	2.33	.07	2.14
FUI45	-.67	2.38	-2.97	1.54	.19	1.33
FUI46	-.78	2.44	-1.94	3.91	-.02	1.39
FUI47	-.49	2.04	-1.84	2.02	.47	1.13
FUI48	-.50	2.40	-1.50	4.16	.41	1.34
FUI49	-1.31	2.88	-2.73	2.22	-.71	2.93
FUI50	-1.24	2.64	-2.05	28.69	-.72	1.81
FUI51	-1.58	2.63			-1.15	1.90
FUI52	-2.02	2.45			-1.66	1.97
FUI53	-1.13	.87	-3.04	.61	-.56	.59
FUI54	-.70	1.60	-1.62	1.61	-.06	1.76
FUI55	-.98	1.50	-2.72	1.22	-.36	.81
FUI56	-1.26	1.41	-2.43	1.30	-.68	1.23
FUI57	-.81	1.60	-3.94	.78	-.02	1.09
FUI58	-.48	1.26	-1.98	1.48	2.91	.17
FUI59	-.70	1.48	-2.84	.82	.02	1.03
FUI60	-1.35	1.74	-1.96	2.30	-.79	1.64
FUI61	-.54	.68	-1.03	.97	.11	.49
FUI62	-.90	1.01	-19.99	.09	-.20	.92
FUI63	.01	1.26	-6.32	.24	1.67	.67
FUI64	-1.04	2.18	-4.12	1.01	-.41	1.68
FUI65	-.86	1.45	-58.03	.05	-.10	1.03
FUI66	-.23	2.51	-1.48	3.45	1.04	1.15
FUI67	-1.66	1.58	-2.04	2.12	-1.06	1.78
FUI68	-.52	1.86	-1.33	2.59	.21	1.46
FUI69	-.72	2.09	-1.65	2.98	.01	1.35
FUI70	-1.97	1.44	-2.27	2.01	-1.46	1.40
FUI71	-.48	1.89	-3.27	.90	.51	1.11
FUI72	-1.30	2.18	-1.41	32.48	-.82	1.67
FUI73	-.51	1.12	-1.57	1.20	.40	.63

Note. Diff=Difficulty; Discrim=Discrimination. FUI items 1, 7, 8, 10, 26, 30, 35, 37, 51, and 52 were removed from the honest sample for these analyses.
[a]N=214. [b]n=101. [c]n=113

Table F3

Factor Scores and Communalities for Combined and Feign Samples on the FUI Based on a Rotated Two-Factor Solution

	Combined			Feign		
	Factor			Factor		
Item	1	2	h^2	1	2	h^2
FUI2	**.62**		.38	**.54**		.29
FUI4	**.44**	.16	.22	**.47**		.20
FUI9	**.75**	.35	.69	**.38**	.32	.39
FUI11	**.61**	.30	.47	**.57**		.29
FUI13	**.63**	.52	.67	**.71**	-.30	.33
FUI16	**.54**		.30	**.43**		.23
FUI17	**.69**	.51	.73	**.79**	-.14	.51
FUI20	**.57**	.33	.43	**.45**	.29	.45
FUI24	**.78**	.19	.65	**.58**		.32
FUI25	**.43**	.28	.26	**.51**		.32
FUI27	**.76**	.33	.69	**.32**	-.19	.06
FUI31	**.64**	.38	.56	**.49**	.31	.52
FUI32	**.61**	.43	.56	**.79**	-.24	.45
FUI34	**.51**	.14	.28	**.54**	.19	.45
FUI36	**.45**	.32	.31	**.50**	.34	.57
FUI37	**.66**	.54	.73	**.57**	.20	.51
FUI38	**.80**	.29	.72	**.75**		.66
FUI41	**.68**	.56	.77	**.76**		.52
FUI42	**.72**	.59	.87	**.48**	.17	.36
FUI43	**.61**	.50	.62	**.85**	-.23	.53
FUI44	**.75**	.52	.83	**.81**		.58
FUI45	**.79**	.38	.77	**.44**	.31	.46
FUI46	**.68**	.49	.69	**.49**	-.10	.19
FUI47	**.79**	.34	.75	**.49**	.28	.48
FUI48	**.77**	.43	.77	**.47**		.23
FUI53	**.50**	.21	.30	**.33**	.25	.28
FUI55	**.55**	.32	.40	**.50**	-.42	.17
FUI58	**.77**		.60	**.91**	-.30	.57
FUI59	**.77**	.22	.64	**.51**	-.12	.20
FUI61	**.45**	.14	.23	**.36**	.20	.25
FUI62	**.45**	.39	.35	**.38**		.18
FUI63	**.62**	.39	.53	**.51**	.23	.46
FUI66	**.73**	.55	.83	**.53**	.24	.49

FUI69	**.64**	.51	.67	FUI71	**.93**	-.31	.60
FUI71	**.84**	.27	.78	FUI72	**.56**	.15	.45
FUI1	.18	**.84**	.74	FUI73	**.30**	.19	.20
FUI5	.18	**.41**	.20	FUI1	-.21	**.97**	.73
FUI6	.23	**.62**	.44	FUI2	.28	**-.37**	.09
FUI7	.44	**.68**	.66	FUI3	.15	**.49**	.35
FUI8	.15	**.80**	.66	FUI5		**.46**	.18
FUI10	.37	**.79**	.76	FUI6	-.11	**.75**	.47
FUI15	.48	**.62**	.62	FUI7		**.72**	.56
FUI18	.50	**.58**	.58	FUI8	-.29	**.98**	.68
FUI19	.53	**.64**	.69	FUI10		**.83**	.68
FUI21	.48	**.54**	.53	FUI12	.19	**.49**	.40
FUI22	.27	**.71**	.57	FUI14	.21	**.68**	.68
FUI23	.34	**.61**	.49	FUI15	.13	**.54**	.39
FUI26	.36	**.72**	.65	FUI18	.21	**.47**	.38
FUI29	.27	**.66**	.50	FUI19	.32	**.49**	.53
FUI30	.41	**.71**	.67	FUI21	.23	**.36**	.29
FUI33	.29	**.74**	.64	FUI22	-.18	**.76**	.44
FUI35	.42	**.72**	.70	FUI23	.18	**.54**	.44
FUI39		**.82**	.68	FUI26		**.74**	.50
FUI40	.13	**.69**	.50	FUI29	-.20	**.68**	.33
FUI49	.38	**.79**	.78	FUI30		**.71**	.54
FUI51	.26	**.83**	.75	FUI33		**.81**	.59
FUI52	.17	**.84**	.73	FUI35		**.73**	.54
FUI56	.43	**.54**	.48	FUI39	-.17	**.96**	.74
FUI60	.35	**.70**	.61	FUI40		**.73**	.46
FUI64	.53	**.62**	.67	FUI49		**.79**	.70
FUI65	.44	**.59**	.54	FUI51	-.25	**1.00**	.75
FUI67	.25	**.71**	.57	FUI52	-.23	**.92**	.63
FUI70	.24	**.65**	.48	FUI56	.25	**.40**	.35
FUI3	**.61**	**.61**	.74	FUI60	.11	**.60**	.45
FUI12	**.55**	**.56**	.62	FUI64	.22	**.56**	.53
FUI14	**.57**	**.54**	.61	FUI65		**.54**	.31
FUI28	**.57**	**.53**	.60	FUI67		**.70**	.46
FUI50	**.56**	**.57**	.64	FUI70		**.60**	.40
FUI54	**.55**	**.53**	.58	FUI13	**.30**	**.27**	.27
FUI57	**.55**	**.50**	.56	FUI68	**.35**	**.39**	.45
FUI68	**.58**	**.55**	.64				
FUI72	**.56**	**.52**	.58				
FUI73	**.48**	**.43**	.41				

Note: Boldface indicates highest factor loadings; FUI items are found in Appendix C. h^2=communality.

References

Alwes, Y., Clark, J., Berry, D., & Granacher, R. (2008). Screening for feigning in a civil forensic setting. *Journal of Clinical and Experimental Neuropsychology, 30,* 133-140.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association.

Bates, T., Davis, T., Guin, C., & Long, S. (1992). Assessment of the literacy levels of adult prisoners. *Journal of Correctional Education, 43,* 172-175.

Bianchini, K., Mathias, C., & Greve, K. (2001). Symptom validity testing: A critical review. *The Clinical Neuropsychologist, 15,* 19-45.

Binder, L.M. (1993). Assessment of malingering after mild head injury with the Portland Digit Recognition Test. *Journal of Clinical and Experimental Neuropsychology, 15,* 170-182.

Borum, R., & Grisso, T. (1995). Psychological test use in criminal forensic evaluations. *Professional Psychology: Research and Practice, 26,* 465-473.

Butchner, J., Dahlstrom, G., Graham, J., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2).* National Computer Systems: Minnetonka, Minnesota.

Colwell, K., Colwell, L., Perry, A., Wasieleski, D., and 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,* 17-42.

Crocker, L., & Algina, J. (1986). *Introduction to modern and classical test theory.* Wadsworth Group: Belmont, CA.

Downing, S. (2006). Twelve steps for effective test development. In S. Downing & T. Haladyna (Ed.), *Handbook of Test Development* (pp. 3-26) Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Dusky v. United States, 362 U.S. 402 (1960)

Everington, C., & Luckasson, R. (1992). *Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST*MR) Test Manual.* Worthington, Ohio: IDS Publishing Corporation.

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

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

Everington, C., Notario-Smull, H., & Horton, M. (2007). Can defendants with mental retardation successfully fake their performance on a test of competence to stand trial? *Behavioral Sciences and the Law, 25,* 545-560.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32,* 221-233.

Frederick, R. (1997). *Validity Indicator Profile Manual.* Minnetonka, MN: NCS Assessments.

Frederick, R. (2000). A personal floor effect strategy to evaluate the validity of performance on memory tests. *Journal of Clinical and Experimental Neuropsychology, 22,* 720-730.

Gothard, S., Rogers, R., & Sewell, K. (1995). Feigning incompetency to stand trial: An investigation of the GCCT. *Law and Human Behavior, 19,* 363-373.

Gothard, S., Viglione, D., Meloy, R., & Sherman, M. (1995). Detection of malingering in competency to stand trial evaluations. *Law and Human Behavior, 19,* 493-505.

Grisso, T. (2003). *Evaluating Competencies: Forensic Assessments and Instruments*, 2nd Ed. New York: Kluwer Academic/Plenum Publishers.

Heckel, R., Allen, S., Andrews, L., Roeder, G., Ryba, P., & Zook, W. (1989). Normative data on the Kagan Matching Familiar Figures Test for adult male incarcerates. *Journal of Clinical Psychology, 45,* 155-160.

Hiscock, C., Branham, J., & Hiscock, M. (1994). Detection of feigned cognitive impairment: The two-alternative forced-choice method compared with selected conventional tests. *Journal of Psychopathology and Behavioral Assessment, 16,* 95-110.

Jacobs, M., Ryba, N. R. & Zapf, P.A. (2008). Assessing competence-related abilities and psychiatric symptoms: An analysis of underlying structure and correlates of the MacCAT-CA and the BPRS. *Law and Human Behavior, 32,* 64-77.

Kerridge, B. (2008). A comparison of U.S. jail inmates and the general U.S. population with Diagnostic and Statistical Manual of Mental Disorders IV alcohol use disorders: sociodemographic and symptom profiles. *Alcohol, 42,* 55-60.

Lipsitt, P., Lelos, D., & McGarry, A. (1971). Competence for trial: A screening instrument. *American Journal of Psychiatry, 128,* 105-109.

Macdonald, P.D.M. and Green, P.E.J. (1988) *User's Guide to Program MIX: An Interactive Program for Fitting Mixtures of Distributions.* ICHTHUS DATA SYSTEMS.

McGarry, A., & Curran, W. (1973). *Competency to Stand Trial and Mental Illness.* Rockville, MD: National Institute of Mental Health.

Melton, G., Petrila, J., Poythress, N., & Slobogin, C. (2007). *Psychological evaluations for the courts: A handbook for mental professionals and lawyers, 3rd Ed.* New York: The Guilford Press.

Miller, H.A. (2001). *MFAST: Miller Forensic Assessment of Symptoms Test professional manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Millis, S., & Volinsky, C. (2001). Assessment of response bias in mild head injury: Beyond malingering tests. *Journal of Clinical and Experimental Neuropsychology, 23,* 809-828.

Morey, L. C. (1991). *The Personality Assessment Inventory (PAI).* Psychological Assessment Resources: Lutz, Florida.

Otto, K., Poythress, N., Nicholson, R., Edens, J., Monahan, J., Bonnie, R., Hoge, S., & Eisenberg, M. (1998). Psychometric properties of the MacArthur competence assessment tool-criminal adjudication. *Psychological Assessment, 10,* 435-443.

Pankratz, L., Fausti, S., & Peed, S. (1975). A forced-choice technique to evaluate deafness in the hysterical or malingering patient. *Journal of Consulting and Clinical Psychology, 43,* 421-422.

Pinals, D., Tillbrook, C., & Mumley, D. (2006). Practical application of the MacArthur Competence Assessment Tool-Criminal Adjudication (MacCAT-CA) in a public sector forensic setting. *The Journal of the American Academy of Psychiatry and the Law, 34,* 179-188.

Poythress, N., Nicholson, R., Otto, R., Edens, J., Bonnie, R., Monahan, J., & Hoge, S. (1999). *The MacArthur Competence Assessment Tool-Criminal Adjudication: Professional manual.* Odessa, FL: Psychological Assessment Resources.

R Development Core Team. (2008). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria.

Rees, L., Tombaugh, T., Gansler, D., & Moczynski, N. (1998). Five validation experiments for the Test of Memory Malingering (TOMM). *Psychological Assessment, 10,* 10-20.

Reicker, L. I. (2008). The ability of reaction time tests to detect simulation: An investigation of contextual effects and criterion scores. *Archives of Clinical Neuropsychology, 23,* 419-431.

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software, 17,* 1-25.

Roesch, R., & Golding, S. (1980). *Competency to Stand Trial.* Chicago, IL: University of Illinois Press.

Roesch, R., Zapf, P., Eaves, D., & Webster, C. (1998). *Fitness Interview Test* (Revised ed.). Burnaby, British Columbia: Mental Health, Law and Policy Institute, Simon Frasier University.

Rogers, R. (1997). *Clinical Assessment of Malingering and Deception, 2nd Ed*. New York: The Guilford Press.

Rogers, R., Bagby, R., & Dickens, S. (1992). *Structured Interview of Reported Symptoms: Professional Manual.* Odessa, FL: Psychological Assessment Resources.

Rogers, R., & Bender, S. (2002). Evaluation of malingering and deception. In A.M. Goldstein & I.B. Weiner (Eds.), *Handbook of Psychology: Forensic Psychology*. New York: John Wiley & Sons, Inc.

Rogers, R., Grandjean, N., Tillbrook, C., Vitacco, M., & Sewell, K. (2001). Recent interview-based measures of competency to stand trial: A critical review augmented with research data. *Behavioral Sciences and the Law, 19,* 503-518.

Rogers, R., Jackson, R., Sewell, K., Tillbrook, C., & Martin. M. (2003). Assessing dimensions of competency to stand trial: Construct validation of the ECST-R. *Assessment, 10,* 344-351.

Rogers, R., Sewell, K., Grandjean, N., & Vitacco, M. (2002). The detection of feigned mental disorders on specific competency measures. *Psychological Assessment, 14,* 177-183.

Rogers, R., Tillbrrok, C., & Sewell, K. (2004). *Evaluation of Competency to Stand Trial-Revised (ECST-R).* Odessa, FL: Psychological Assessment Resources.

Shipley, W. C. (1940). A self-administered scale for measuring intellectual impairment and deterioration. *The Journal of Psychology, 9,* 371-377.

Shipley, W. C. (1940). A self-administered scale for measuring intellectual impairment and deterioration. *The Journal of Psychology, 9,* 371-377.

Skeem, J., Golding, S., Cohn, N., & Berge, G. (1998). Logic and reliability of evaluations of competence to stand trial. *Law and Human Behavior, 22,* 519-547.

Tombaugh, T. (1996). *TOMM: The Test of Memory Malingering*. North Tonawanda, NY: Multi-Health Systems.

Tombaugh, T. N. (2002). The test of memory malingering (TOMM) in forensic psychology. *Journal of Forensic Neuropsychology, 2,* 69-96.

Vitacco, M., Rogers, R., Gabel, J., & Munizza, J. (2007). An evaluation of malingering screens with competency to stand trial patients: A known-groups comparison. *Law and Human Behavior, 31,* 249-260.

Vitacco, M., Rogers, R., & Gabel, J. (2009). An investigation of the ECST-R in male pretrial patients. *Assessment, 16,* 249-257.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale-4th Edition*. San Antonio, TX: Pearson Assessment.