Fitness/Competency to Stand Trial: A Conceptual Overview, Review of Existing Instruments, and Cross-Validation of the Nussbaum Fitness Questionnaire

David Nussbaum, PhD
Megan Hancock, BSc (Honors)
Ian Turner, BA (Honors)
John Arrowood, PhD
Sheldon Melodick, EdD

Competency to Stand Trial or Fitness to Stand Trial (FST) is the most frequent referral issue facing forensic mental health professionals (FMHPs) and consumes considerable scarce resources in the process. This article summarizes minimalist and expanded legal approaches to FST and briefly describes three instruments developed by FMHPs to structure FST assessments. We then present evidence supporting the validity of the Nussbaum Fitness Questionnaire for efficiently screening individuals for fitness and blatant or subtle malingering. The paper ends with a number of suggestions to optimize use of these instruments within the current set of forensic mental health practices. Specifically, it is suggested that use of the screening instrument could reliably eliminate up to 70% of current referrals for complete assessments while the more in-depth semistructured interviews be utilized to confirm unfitness, especially when the mental health professional has more than trivial doubt regarding an individual's FST. [Brief Treatment and Crisis Intervention 8:43–72 (2008)]
Not surprisingly, CST assessments are the most frequent targets for referral questions asked of forensic experts (Rogers & Mitchell, 1991), exceeding 50,000 annually in the United States. This workload, along with the desire to provide fundamental fairness, has spawned two camps with respect to the issues, reflecting pragmatic and ideological considerations. Dwindling forensic mental health resources increases the value of efficiency. Conversely, if quality were sacrificed for expediency, the complete venture would appear to be for naught. Requirements for evidence-based practice make it incumbent on forensic experts to combine utility with empirically demonstrated validity.

This article will provide the legal background to the construct of FST/CST, describing different legal standards, outlining methods employed to conduct these assessments, examining the recent corpus of research in the area, and empirically demonstrating a number of FST techniques that satisfy these goals. The overall discussion will then comment on the current state of the legal construct and the utility of this class of instruments and will make suggestions for how mental health professionals might optimize use of these instruments.

Legal Background

The Western approach to criminal justice is centered on a few basic principles that underlie the myriad statutes, legal decisions, and less formal rules of conduct that comprise the corpus of criminal law and its practice. One basic tenet of Western criminal jurisprudence is that an individual is presumed innocent until proven guilty and understands the charges. Similarly, there has been a realization in Judeo-Christian thinking going back to ancient times that guilty acts need not necessarily emanate from guilty minds (Babylonian Talmud, Tractate Moed Katan, 3b) rendering punishment in such cases unjustifiable. A similar reluctance to wrongly punish an innocent individual applies to cases where, on account of a mental disorder or limitation, an individual is unable to respond effectively to his or her accusers. Being that an impaired defendant is a vulnerable defendant, fundamental justice dictates an assurance that those standing accused before the courts possess sufficient ability to at least assist counsel and thereby participate in mounting a defense at trial.

There exists a spectrum of thought concerning “how competent is competent.” We will dub one end of the spectrum the “purists” who would require that each defendant be optimally equipped to mount a “best possible” defense. Considerations of court expediency and assessment resources are relegated to secondary importance. Alternately, the “pragmatists,” while recognizing the pitfalls of having an incompetent individual stand trial, wish to ensure that each defendant possesses the necessary ability to be adequately defended but place equal value on efficiency within the judicial system. This philosophical difference manifests itself at both the statute and evaluation levels. It should be noted that the terms purist and pragmatist are used descriptively, not pejoratively, and that both positions have merit. Most practitioners fall somewhere between these extreme positions.

Legal Standards

The legal standard of FST was established to ensure that mentally disordered defendants in criminal trials are cognitively capable of assisting their attorneys in their legal defense (Rogers & Mitchell, 1991). Formally, the legal criteria for deciding fitness in Canada are as follows: An individual is presumed fit to stand trial unless, on account of mental disorder, the defendant is incapable of (a) understanding the nature or object of the proceedings, (b) understanding the
possible consequences of the proceedings, and (c) communicating with counsel (Criminal Code of Canada, 1993, Section 72).

Similar criteria are found in other jurisdictions. In the United States, in Dusky (1960), the Supreme Court rejected the notion that passing a simple factual knowledge test demonstrates sufficient proficiency but established a more restrictive CST standard. The Dusky ruling mandated that the defendant understand the charge, legal proceedings against him/her and be able to assist in his or her defense. Furthermore, the defendant’s knowledge does not have to be comprehensive but rather sufficient and can be taught in the case of ignorance. Competency refers to the current period of legal proceedings rather than some time in the past, as is the case for insanity evaluations. Finally, the Dusky criteria assume a passive defendant who is simply able to “assist” his attorney.

Justice Bonnie (1992) went beyond the Dusky (1960) framework, structuring competency into two streams, (a) ability to assist counsel and (b) decisional competence. Assisting counsel includes (a) understanding of the charge and basic operation of the criminal justice system, especially the role of the defense counsel, (b) appreciation of one’s situation as a defendant, and (c) ability to relate important information to the defense attorney. Decisional competence typically involves competency to waive constitutional rights, such as the right to a trial and to an attorney. Bonnie argued that the level of competency required of a defendant depends on a variety of factors, including the defendant’s cooperation with the attorney and the protections being waived. Consequently, a defendant pleading guilty against legal advice might be required to demonstrate higher levels of competence than a defendant not waiving those rights but accepting guidance. Some American states and the Criminal Code of Canada have not endorsed Bonnie’s inclusion of decisional competence.

Rogers (2001) recently summarized three competing models of CST derived from the Dusky (1960) standard. The first of these was the Discrete Abilities Model, designed to operationalize each element of Dusky. This results in (a) rational ability to consult, (b) factually understanding the proceedings, and (c) rational understanding of the proceedings. Clinical researchers prefer this model (Rogers, Jackson, Sewell, Tillbrook, & Martin, 2003) as it clearly identifies and can quantify separate components of the model. The Domain Model categorization is based on the hyphenation between (a) rational ability to consult and (b) factual and rational understanding of the proceedings. A number of legal scholars favor this model (Rogers et al., 2003). Its main advantage is its separation of competency into only two related domains. The Cognitive-Complexity Model innovatively distinguishes two components of competency based on distinct cognitive abilities: (a) factual understanding and (b) rational abilities. Factual understanding typically entails the simple recall of “crystallized” knowledge about the legal system. Rational abilities on the other hand often require nuanced cognitive processes involving analyses of available information into basic entities, integration of these entities into alternate constructs, and then, arriving at soundly reasoned decisions in light of consideration of complex options. Furthermore, factual understanding appears less vulnerable to the effects of severe psychopathology (Rogers et al., 2003). The main advantage of this model is its sophistication in appreciating the differing cognitive capacities attributed to the Dusky standard.

We favor the minimalist position for theoretical and practical reasons. The more demanding decision and rational standards appear unrealistically high and well beyond what would realistically be needed to comprehend most cases involving the mentally ill. These standards might not be met by individuals with IQs in
the low 80s that are not mentally ill, although to the best of our knowledge, this remains an open empirical question. The accused’s attorney is generally available to make the sufficiently abstract points comprehensible so that the mentally disordered client can make an informed choice regarding a defense. The minimal standard would also appear to be more practical in facilitating trials. Should attorneys be unable to receive instruction from clients who have slipped through the cracks, they have the opportunity to inform the court and fashion a fitness assessment at any point in the trial. Additionally, a trial can proceed with the judge issuing explicit orders to defense counsel to periodically update the accused as to the progress and tack the trial is taking.

**Psychological Considerations and the Competency Construct**

A number of general but still compelling points regarding the role, scope, and methodology necessary to maximize the appropriate contribution of the mental health professional to the courts can be traced to Grisso’s (1986) classic work. These include the nonequality of a mental disorder (some schizophrenics may be competent to stand trial while others are not) and the nonequivalence of different specific legal competencies, that is, competence to stand trial does not imply competency to parent, or criminal responsibility. Finally, there is the necessity to develop specific forensic assessment instruments to reflect the particular skill and knowledge domains demanded by particular legal tests. Developed forensic instruments therefore must mirror “legally relevant functional abilities” (Grisso, 1986, p. 40). Grisso argues that explicit instruments will prevent forensic mental health professionals (FMHPs) from committing the three “cardinal sins”: (a) ignorance of and irrelevance to legal issues, (b) intrusion of personal values and attitudes into legal decision making, and (c) insufficiency and incredibility of proffered information (Grisso, 1986, p. 8). One essential consideration is that the trier of fact is responsible for the ultimate decision and the mental health professional is there to assist in setting the table provide a “penultimate decision” rather than to assume the judiciary’s role in pronouncing the “ultimate” decision. Another key point mentioned by Grisso (1986, p. 20) is the causal link between the identified disorder and the specific incompetence. The dearth of neuropsychological validation in the CST domain is consequently surprising and will be discussed below.

**Recent Psychological and Psychiatric Contributions to the CST Assessment Process**

Gutheil (1999) has argued that psychiatrists and the public share a confused sense of forensic terms including competence and insanity and has written a basic review of the constructs underlying each. That article is appropriate for someone wishing to gain an overview of the areas. Gutheil focuses on the defendant’s factual and rational capacities and ability to communicate with counsel but not on decisional capacity.

Skeem, Golding, Cohn, and Berge (1998) have argued that the trier of fact’s assignment of weight to a CST opinion should reflect the appropriateness of the mental health professional’s understanding of the relevant CST legal criteria and the reasoning underlying the connection between a diagnosed mental disorder and the incapacity said to compromise that ability. In reviewing randomly selected CST reports, they found adequately documented clinical findings but deficiencies regarding links between legally relevant criteria and CST opinion. Despite good overall agreement in terms of CST decision, per se, considerable
divergence existed regarding the reasons for these decisions.

Veiel and Coles (1999) operationalized the CST construct differently. They argue, quite reasonably, that fitness is a relative concept as it must be defined individually relative to the demands of a particular trial. They claim, “there is no readily identifiable biological, psychological or psychiatric referent” and insist that “it is a foreign concept for which there is no direct translation” (Veiel & Coles, p. 357). They argue that fitness cannot be identified but only falsified and is defined by exclusion, it is not a unitary but a synthetic concept composed of a number of unrelated capacities. “It requires the expert to assess the functional abilities that contribute to fitness/unfitness rather than assess fitness/unfitness per se” (Veiel & Coles, p. 357). Mental health practitioners err when they infer the trait of fitness/unfitness from symptoms of mental disorder. This is based partially on the idea that traits exist independent of situations while competency is situation specific.

We now move from the theoretical to evaluation practices. One could argue that because the behaviors typically accompanying incompetency/unfitness are well defined, precise, and readily observable, psychological measurement procedures designed to contend with measurement error are unnecessary. Further, because psychometric instruments are to be interpreted probabilistically while fitness is a deterministic issue (i.e., either fit or unfit), probabilistic measurement approaches are “inappropriate.” They proceed to criticize the development of their own instrument (Coles, Freitas, & Tweed, 1996) for assessing fitness in individuals suffering from mental retardation as it reflects a unitary focus on understanding. They comment on the lack of attention paid to the specific deficits of mentally retarded individuals in that they often learn to reply with stock answers to facilitate social interactions but have very little understanding of the concepts that they articulate.

Veiel and Coles (1999) conclude that psychological measurement is inappropriate for CST and, at most, the roles for psychological instruments should be limited to construct explication and standardization and validation of procedures for component constructs. They favor a branched questioning model with each successful answer entailing a question designed to establish the limits of comprehension of the previous answer, proceeding to asymptote. The purist might argue that Veiel and Coles present a “purist” approach to CST but in a nihilistic fashion. It is not clear how one might better accomplish the branching set of items than recent approaches to be described below in a traditional psychometric fashion. Finally, given that the ultimate decision regarding CST is the task of the trier of fact, a probabilistic statement may be very valuable to her or him in arriving at the ultimate decision in a fashion most physicians report on possible treatment outcomes to patients.

**Neuropsychological Correlates of CST**

Neuropsychological tests contain a unique source of data that may provide information on specific cognitive processes that are critical to CST (Grisso, 1986). Use of these tests may be extended beyond their general clinical applications and may explore a relationship between competency and the ability to acquire, encode, retain, and retrieve pertinent legal information (Nestor, Daggett, Haycock, & Price, 1999). A study conducted by Nestor et al. (1999) identified distinct cognitive capacities of defendants charged with serious felonies and recommended as either competent to stand trial (CST) or incompetent to stand trial (IST) by applying standardized neuropsychological measures. The cognitive functions included episodic
memory, semantic knowledge, and social intelligence. This study demonstrated that a clinical neuropsychological information-processing model could explain cognitive dynamics of competence.

The data set included 309 patients and examined assessments completed between 1987 and 1995. Intelligence was measured by the WASI-R, memory by the WMS-R. Measures of attention and concentration included the Trail Making Tests and the WMS-R Attention and Concentration Index. Academic abilities were indexed by the WRAT-R. Finally, the Wisconsin Card Sort Test was employed for problem solving and executive functions abilities (Nestor et al., 1999).

Psychometric measures of intelligence and memory were strongly related to the competence-related deficits that resulted in clinical IST judgments. On the other hand, language-related academic abilities of reading, spelling, and written arithmetic were not strongly related to the clinical judgments of competence. While the IST group scored more poorly than the CST group, these differences failed to reach statistical significance (Nestor et al., 1999). These results suggest that fit and unfit groups differ with respect to IQ and memory intactness but not to crystallized (and possibly static) factual knowledge.

Similarly, Kirkish and Sreenivasan (1999) argue that neuropsychological test data can be applied to CST evaluations. These authors contend that competency assumes a number of basic abilities such as receptive language skills, expressive language skills, memory, attention, and executive functions. Neuropsychological assessments evaluate these abilities and ensure that they are intact. Additionally, when neuropsychological impairment is raised as an issue in CST evaluations, the extent and nature of deficits and whether the deficits are representative of a genuine or malingered neurobehavioral disorder can be assessed via a neuropsychological assessment. The authors argue that if an association between an observed cognitive deficit and impairment in one of the legs of CST is discovered, an opinion of incompetence can be rendered with appropriate evidence and enhanced confidence.

In summary, because neuropsychological tests measure constructs such as attention, comprehension, and memory, they provide an independent domain-free reflection of an accused’s ability to participate meaningfully at trial and should be considered useful adjuncts to legally focused CST evaluations as well as important measures of construct validity in research studies.

**Recent Research**

To help organize the research, we will follow the natural course of CST assessments.

**Who Should Be Referred?** In most jurisdictions, the issue of CST can be raised by any of the three parties (judiciary, prosecution, or defense) at any preverdict phase of the trial. Largely unstudied is the nature of who is sent for an assessment. It might be inferred that different jurisdictions have different thresholds for who is referred, as Skeem et al. (1998) report 53% of the decisions in their study as incompetent as compared to a 25%–30% average reported in an earlier review by Nicholson and Kugler (1991). Earlier, Grisso (1986) reported ranges between 4% and 77%. Skeem et al. note that these differences do not reflect interrater variance because the approximately 50% was consistent across raters in their study. To the extent that referrals are limited to the most severely disordered individuals, the rate of unfit will increase, although the overall number of referrals will be low. Skeem et al. cite the finding of Maguire and Pastore (1994) that in the United States during 1993, 49,611 defendants (approximately 2% of all criminal
cases) were referred for a CST evaluation, resulting in a cost of $467 million 1993 American dollars. Clearly, a valid screening device would have great utility that both pragmatists and purists would welcome if it could effectively identify those who are found competent in subsequent hearings without missing those found incompetent.

Cooper and Grisso (1997) noted the relative decline in research concerning the extent to which competency assessments have become a “back door” for the families, police, and courts, often acting in arguably benign collusion, to obtain treatment for recalcitrant mentally ill individuals. Based on cited research, Cooper and Grisso conclude that most of the competency referrals would either be considered dangerous, and thereby qualify for civil commitment (Appelbaum, Fisher, Nestelbaum, & Bateman, 1992), or occur in jurisdictions where there has been a reduction in mental health beds (Geller, Fisher, & Kaye, 1991). This has been referred to as the “poor man’s mental health system” by Schnieder (1996). To the extent that various jurisdictions have implemented successful “diversion programs” intended to channel “nondangerous” mentally disordered individuals into the appropriate mental health system, this issue is perhaps less salient than when the review was written by Cooper and Grisso (1997).

Who Can Perform CST Assessments? The survey of 55 American states and territories by Farkas, DeLeon, and Newman (1997) found that all 52 responding jurisdictions sanctioned psychiatrists to do these evaluations. Forty-two jurisdictions did not require any specialized certification beyond the professional license. Ten states required additional forensic certification all of which also sanction psychologists for CST evaluations. Six of these states qualify social workers, four nonspecialist MDs, and three qualify “others.” Of the 37 states that do not require forensic certification, only five (Delaware, Iowa, New Jersey, Pennsylvania, and Vermont) do not sanction psychologists to testify in CST cases. Unfortunately, the Criminal Code of Canada does not mention psychologists within the “who may” paragraph but does require a physician (and not necessarily a psychiatrist) to perform these assessments.

Where Are CST Assessments Performed Optimally? Prior to Allan Guttmacher’s groundbreaking establishment of the Office of the Medical Examiners in Baltimore, CST assessments were typically conducted in jail cells. The local guards provided security for the examining professional. This practice could threaten the integrity of the CST decision, although no research has addressed this because often even competent defendants such as child molesters may not wish to disclose the nature of their alleged offense to someone who may carry that knowledge back to the cell block. It could be argued that jail cells provided less than an optimal milieu in which to establish even basic rapport because the examiner was perceived to be part of the punitive, adversarial, and oppressive judicial system. Consequently, some jurisdictions such as Ontario (with the establishment of the Metropolitan Toronto Forensic Service [METFORS]) established off-site hospital units to provide multidisciplinary teams to perform these assessments, but these self-contained court clinics involve considerable costs in terms of transportation and extra staffing. For example, Chaimowitz and Ferencz (1999) developed a “pilot fitness clinic” in a detention center in Hamilton, Ontario, that was attached to the local Hamilton Psychiatric Hospital Forensic Program. The contingent consisted of a psychiatrist, social worker or nurse, and student observer. They estimated an average of 5.25 hr of clinician’s time per assessment, with interviews lasting between 10 and 60 min. They note improvements in
wait-list time between court order and the detention center assessments compared to previous waiting times. Seventy-four percent of referrals were found competent, consistent with the finding of Roesch, Ogloff, Hart, and Dempster (1997) that 79.4% were competent. Overall, they estimate that their procedure saved $400,000 Canadian (or about $360,000 US at the time of this writing) annually for the catchment area of 1.5 million people serviced by a 446-bed psychiatric facility.

Chaimowitz and Ferencz (1999) did not employ a random assignment and a control group to establish the potential validity of evaluations conducted in detention versus hospital sites. To the extent that factors such as reluctance to disclose and a perceived hostile environment influence the ease with which defendants disclose, purists might contend that such considerations mitigate against CST assessments being conducted in detention centers and jails despite their clear advantages with respect to cost saving, shortened waiting periods, and convenience.

What Should Be Considered in a CST Assessment and Report? Skeem et al. (1998) conducted a study examining two CST reports on each of 50 defendants paying specific attention to examiner’s (a) explanations and support of the defendant’s CST abilities, (b) explanations and support of the defendant’s symptoms of psychopathology, (c) ability to provide data and reasoning linking these symptoms to psycholegal deficits in support of CST conclusions, (d) agreement of opinions with third party sources of information and assessment instrument data, and (e) concordance with another examiner’s judgment of the defendant’s clinical status and CST and on the support for these opinions.

Skeem et al. (1998) provided a number of criticisms and suggestions regarding CST assessments and reports. Firstly, it was discovered that while examiners normally provided adequate reasoning to substantiate their clinical conclusions, they provided inadequate reasoning to support their psycholegal conclusions. As a result, little or no reasoning addressing the correlation between the clinical description of the defendant’s symptoms and impairments in CST abilities was included in the reports. Reports should explicitly inform the court about the nature of the relationship between the defendant’s psychopathology and deficits in psycholegal abilities, as legal criteria consistently link mental disorder with competence deficits (Skeem et al., 1998).

Examiners tended to concur as to defendant’s global diagnostic category (79%) and competency (82%); however, this was lower than the rates of agreement (e.g., 90%–97%) when structured CST assessment instruments in joint or independent interviews were used by forensically trained examiners (Skeem et al., 1998). In addition, examiners infrequently presented similar support for their opinions, predominantly with respect to specific CST deficits. It is critical that reports provide data and reasoning to substantiate the defendant’s specific psycholegal impairments and abilities as judges comparatively evaluate the reasoning expressed in the reports to render an independent judgment (Skeem et al., 1998).

Finally, while the arrest report was reviewed by most examiners, most reports did not describe contacting the defendant’s attorney which is a contact that could shed light on relevant areas of question by identifying the reason for the referral and the likely demands of the defendant’s case (Skeem et al., 1998). In addition, a minority of examiners reviewed defendants’ mental health records, an act that may assist in detecting malingered deficits. Consulting with such sources of information is a fundamental part of completing a relevant, informed, and valid forensic assessment in addition to preparing the examiner for an adversarial attack (Skeem et al., 1998).
Recent Instruments Designed to Assist CST Evaluations

In this section, we will describe four instruments designed specifically to help in the assessment of CST. For each, we will describe the legal standard addressed, format, number of evaluative empirical studies in the published literature, a description of those empirical findings including reliability coefficients, validation statistics (decision convergence with concurrent measures and the adequacy of those measures), other psychometric properties, and unique features where appropriate, and an evaluative summary. We briefly review three instruments in common use and then present two empirical studies validating the METFORS Fitness Questionnaire (MFQ, Nussbaum, Mamak, Tremblay, Wright, & Callaghan, 1998, renamed the Nussbaum Fitness Questionnaire [NFQ] in its current form). The reviewed instruments include (a) the Fitness Interview Test (Revised edition) (FIT-R, Roesch, Zapf, Eaves, & Webster, 1998), (b) the MacArthur Competence Assessment Tool-Criminal Adjudication (MacCAT-CA; Hoge, Bonnie, Poythress, & Monahan, 1999), and (c) the Evaluation of Competency to Stand Trial-Revised (ECST-R; Rogers et al., 2003).

The Fitness Interview Test (Revised Edition) The FIT-R, a refinement of the earlier FIT (Roesch, Webster, & Eaves, 1984) provides a semistructured interview approach to conducting FST evaluations within the Canadian legal context. The FIT-R was intended as a screening instrument (favoring overestimates of unfitness). The manual outlines a study by Zapf and Roesch (1997) that found perfect sensitivity for the FIT referenced against the regular institutional (psychiatric) opinions and a 14% false positive error rate in 57 assesses. The manual describes the format of the FIT-R within three sections focusing on (a) factual knowledge of criminal proceedings (i.e., nature and object of the proceedings), (b) appreciation of personal involvement and importance of consequences, and (c) capacity to communicate with counsel and participate in defense.

The semistructured interview begins with four “background” questions regarding legal representation. Section 1 addresses “understanding of nature and object of the proceedings” querying knowledge of items such as the arrest process, severity of charge, and roles of key participants at trial. Section 2 concerns “appreciation of personal involvement and important consequences” including appraising knowledge of legal defenses and likely outcome. Section 3 evaluates “capacity to communicate with counsel and participate in defense” and consists of items related to a variety of communicative capacities such as relate the facts of the case to the lawyer, challenge prosecution witnesses, and testify relevantly.

These different subdomains are collectively rated on a three-point scale (0–2 ranging from no to severe impairment. Opinion regarding Fitness rests on the basis of an existing legally recognized mental disorder and a related inability to perform any of the three competency-related functions noted above decided on the basis of global “yes,” “possibly,” and “no” ratings on the relevant items. There is no “cut score” calculated on the individual items within each domain to assist with translating the pattern into one’s overall penultimate decision.

Research on the FIT-R. Given that the FIT-R is a relatively new test, there is an appropriate body of research on this instrument beyond that provided in the manual. Viljoen, Roesch, and Zapf (2002) conducted a study in which physicians, forensic psychologists, nurses, and graduate students in psychology were trained on the FIT-R. They all received standardized instructions and viewed videotape on the FIT-R. Following this training, the
participants viewed videotapes of actual fitness assessments, and using the FIT-R they rated the fitness of each defendant. It was found that highly reliable screening decisions could be made by the evaluators about the defendants’ overall ability to stand trial. The average intraclass correlation was .98 for the overall judgment of fitness, and reliability estimates were high across professional groups. The results of this study provide additional support for the psychometric properties of the FIT-R, as well as for the ability to conduct reliable fitness assessments using this measure by a variety of professionals.

A more recent study by Viljoen, Vincent, and Roesch (2006), interested in identifying assessment tools that could be used with juvenile defendants, evaluated the properties of the FIT-R for that purpose. More specifically, these researchers examined the interrater reliability for items, sections, and overall determinations of competence and compared the interrater reliability of structured clinical ratings to numerical summary scores. Furthermore, the structural validity of the FIT-R was investigated using confirmatory factor analysis to test whether the FIT-R adhered to its theoretical three-factor discrete abilities structure (understanding, appreciation, and the ability to communicate with counsel).

The authors interpreted these results to conclude that the interrater reliability of the FIT-R was satisfactory and its factor structure was relatively consistent with its rationale and organization when used with a sample of juvenile defendants (Viljoen et al., 2006).

Critique. The FIT-R has much to recommend its use. Although the authors note that it may be used as a screen, its thoroughness makes one wonder what else might be clarified regarding competency, even during an inpatient stay predicated on an inconclusive initial interview. Clinical experience suggests that for individuals experiencing acute psychotic episodes, the rigor of the FIT-R interview may exceed their attentional capacities and focus. A series of partial administrations might more reliably tap impaired patients’ attentional capacities although fail to replicate durations typical criminal trials.

One must be somewhat cautious in using the FIT-R (or any other instrument) as a sole arbiter of competency without the assessor having a thorough understanding of the statutory requirements in one’s particular jurisdiction and the “fit” between the instrument and relevant legislation. For example, the Taylor (1992) standard used across Canada, as noted above, is a limited cognitive capacity test. There are a number of examples of specific FIT-R items that may go beyond the Taylor criteria in that they appear to exceed the minimal threshold envisioned in Taylor (and possibly some interpretations of Dusky, 1960). These items involve insight into areas such as assuming other’s perspectives, specialized knowledge of the appropriate sections of the criminal code, specialized knowledge of the federal state or provincial penal system, the practice of criminal law, as well as being able to understand the particular judge or jury’s views concerning the crime type and similar circumstances.

For these reasons, it is likely that anyone found competent when employing the FIT will in fact be so but the issue lies in the ability of the FIT (and other tests) to distinguish those actually competent who might not appear so on initial examination. To the extent that the FIT speaks to a relatively high standard, it may declare some individuals incompetent who actually have the wherewithal to meaningfully participate in their trial, especially in relatively simple and straightforward cases.

Operationally, the FIT provides a general description of each capacity under consideration and lists a number of questions to probe the accused’s knowledge or ability in particular domains As with many structured and
semistructured approaches to forensic assessment, while the added organization greatly enhances uniformity, consideration of nonobvious domains, objectivity, and rigor to the assessments, an objective algorithm for relating the elicited responses is not provided for the FIT. There is no suggested reference to objective impairment scores on standardized psychotic symptom scales or neuropsychological testing of sustained attention or memory. Legal competencies are actually moral/social constructs and not scientific propositions with objective gold standards (Grisso, 1986). Nevertheless, the use of empirically validated measures of impairment severity might be useful to ground subjective ratings in time-constrained contexts.

Despite these minor points, the FIT-R has excellent psychometric properties in terms of interrater reliability and (as will be described below) concurrent validity with the MacCAT-CA and judicial decisions. The FIT-R is clearly a purist model and any accused passing its rigorous criteria would convincingly be found competent by a trier of fact. The empirical studies support its use and ability to detect the unfit as well.

The MacArthur Competence Assessment Tool-Criminal Adjudication. The MacCAT-CA was developed as part of the MacArthur Foundation Research Network on Mental Health and the Law initiative. Earlier (first generation) CST instruments were designed to provide standardization to enhance clinicians’ efforts in assessing CST (see above). Although the MacCAT-CA has been reviewed by Melton, Petrilla, Poythress, & Slobogin [1997], their review preceded publication and subsequent research. It is consequently reviewed in some detail here.) As noted, these generally focused on defendant’s understanding of the legal system (CST and GCCST) or included a wider range of constructs for consideration but do not provide a uniform set of questions, objective scoring system, or norm referenced interpretive guidelines (CAI and IFI; Poythress et al., 1999). Additionally, although Dusky had been in place since 1960, the US Supreme Court decision in Godinez v. Moran, 1993, did recognize a second component, decisional competence, to abilities required for autonomous decision making regarding strategic issues arising in the course of a trial. Poythress et al. (1999) state that this additional requisite subsumes:

... the abilities to seek out and weigh rationally the facts or data that inform defendants’ choices on legal issues in which the legal system demands a degree of autonomous participation (e.g., whether to testify; whether to plead guilty or not guilty and proceed to trial.) (p. 2).

The so-called first-generation instruments, developed prior to 1992, did not have to address these decisional competencies. The MacCAT-CA represents a psychometrically informed and pragmatic distillation of a lengthier precursor. Nevertheless, there is sufficient scope and rigor to satisfy the most dedicated purist.

The MacCAT-CA is an innovative, thorough, and realistically objective instrument for performing CST evaluations. It is concerned explicitly with ecological validity in relating to the accused’s particular incident in addition to reflecting on one’s ability to provide answers to hypothetical legal questions. The vehicles by which the MacCAT-CA accomplishes this task is by providing an incident and then asking a series of questions regarding the incident and extensions of the process. The incident itself involves two men who play pool in a bar and engage in a fight. One hits the other with a pool cue; the victim falls to the floor, and strikes his head so hard that he nearly dies.
The questions regarding CST are then put to the accused in a logical sequence. Three ability areas are tapped including understanding, reasoning, and appreciation. Items are scores on a 0, 1, 2 basis similar to the FIT-R. Unlike the FIT-R, MacCAT-CA norms and percentile rankings are provided for each of the three abilities. Additionally, clinical observations are noted relating to “ability to remember relevant events,” “ability to communicate in a coherent manner,” “ability to function in courtroom roles,” and “other.” Possible impairment or the absence thereof is then noted and the clinician makes appropriate use of such observations.

Another innovative feature of the MacCAT-CA relates to what is commonly referred to as “competency (or fitness) teaching.” In the course of CST evaluations, the professional is often left to wonder whether an individual’s wrong or absent response is due to the important distinction between a defendant’s mental disorder or legal culture ignorance. To provide an opportunity for the accused to answer an item correctly and achieve competency, provision of correct answers are provided when necessary and, at a later time, the individual is reexamined to determine whether he/she has assimilated the dispensed information. Typically, this is done in an informal fashion. To accomplish this more systematically, the MacCAT-CA has instituted a second run called the “disclosure” whereby if an individual receives either a partial credit or no credit for an item, he/she is immediately provided with a specific encapsulation of the relevant information and is then rescored on the item. If the individual can provide the correct response to the examiner in a nonparroted fashion, the higher score from the disclosure will be counted. If the disclosure does not result in an improvement, the original item score stands. Thus, the accused’s ability to learn the relevant information is assessed immediately in a standardized fashion, and this is taken into account in arriving at an opinion regarding competency.

Scoring criteria are conveniently listed along the facing page of the questions, which has space for rescoring the answer verbatim as well as item a scoring slot. In the event of failure or incomplete success, the disclosure question is displayed near the bottom of the page with room again for the response and scoring key. The scoring criteria in the Interview Booklet and the Manual are explicit and unambiguous. The Manual contains four helpful chapters, including a chapter discussing the concept legal competencies and a review of earlier research; a chapter on appropriate assessment populations, administration, and scoring instructions and how the MacCat-CA is to be used within the framework of the overall competency evaluation; a chapter on test interpretation with case study examples; and finally a chapter on psychometric properties.

Content of MacCAT-CA Items. Understanding items. Under “understanding,” test items relate to recognition of the roles of the defense and prosecuting attorneys, with additional items including the requisites for aggravated and simple assault, jury properties, judges’ functions, available pleas, burden of proof, implications of a guilty plea, sentencing guidelines, right to offer evidence of innocence and challenge allegations, right to remain silent/protection against self-incrimination, and the possibility of appeal.

Reasoning items. The first reasoning item, (Item 9) relates to recognition from the vignette that one of the presented facts is more helpful than the other in constructing a self-defense defense. Additional items involve selecting facts illustrating a lack of intent to injure, suggesting provocation, recognition of the defendant’s need to protect himself, recognition of alcohol’s
potential to impair judgment and behavior, the ability of accused to solicit new and potentially heuristic information about the case, and the ability of the accused to see both implicit advantages and disadvantages of pleading guilty or not guilty. The final reasoning item requires that the accused provide at least two reasons why the chosen alternative (or a guilty plea if no choice is made) is superior to the alternate choice (or no choice).

Appreciation items. Unlike the previous two item clusters, the appreciation items all relate to the particulars of the defendant’s trial. Consequently, accurate scoring requires that the evaluator has detailed information about the defendant’s criminal and mental health histories, index offense, capability of the opposing attorneys, and the judge’s track record in cases involving similar offenses and defendants. Test items address the accused’s ability to evaluate the likelihood of his/her being treated fairly by the court, the likelihood of being helped by defense counsel, the likelihood that the individual will fully disclose case material to his/her attorney, and the defendant’s ability to estimate the likelihood of his/her being found guilty. Additional items canvass one’s ability to estimate likelihood of punishment if convicted and the likelihood of the defendant pleading guilty.

Psychometric Characteristics. A detailed account of the complete development and norms for the MacCAT-CA can be found in Otto et al. (1998) and in the Manual. The total sample consisted of 729 individuals derived from three groups: (a) 197 unscreened jail (jail unscreened or JU) inmates, (b) 249 treated jail inmates (i.e., receiving mental health services; JT), and (c) 283 hospitalized incompetent (HI) defendants drawn from a variety of settings in eight states to help ensure representativeness of the sample. Individuals with estimated IQs below 60 (based on WAIS-R Information and Picture Completion) as well as those suffering from cognitive disorders were excluded from the study. The authors presumed that the two jail samples were competent while the HI group, evaluated within 2 weeks of their hospitalization, remained incompetent. A minor point is that such presumptions, especially with regard to the JU group, would likely depress the group norms somewhat because the cut scores will reflect data in part from incompetent individuals treated as if they were competent. The groups did not differ on demographic variables.

Importantly, Hoge et al. (1999) did use psychometric instruments to measure impairment in evaluating validity of the MacCAT. Not surprisingly, different subgroups (with respect to competency) differed on all Brief Psychiatric Rating Scales, most notably the Psychotocism scale. The different subgroups did not differ on estimated IQ, but this is likely attributable to the choice of subtests used, which is likely more reflective of premorbid IQ than current intellectual ability. Were current IQ in its broad manifestation (which includes at very least understanding and reasoning ability) actually equivalent between the groups, the construct validity of the MacCAT-CA if not the competency construct itself could be seriously questioned.

MacCAT-CA Reliability and Validity. Cronbach’s alpha (reflecting internal consistency of items) was high (.85, .81, and .88 for understanding, reasoning, and appreciation items, respectively). Interrater reliability sampling from eight raters was also high with overall intraclass correlations for understanding, reasoning, and appreciation being .90, .85, and .75, respectively.

Hoge et al. (1999) evaluated a range of validity measures in establishing the construct validity of the MacCAT-CA. Separate norms and score ranges for minimal/no impairment, mild impairment, and clinically significant
impairment are presented for understanding, reasoning, and appreciation as interpretation guidelines. These three CST domain scores statistically differed between the HI and the two jailed groups although the JU and JT groups did not differ. Moderate correlations speaking to concurrent validity were found between IQ estimate and both understanding and reasoning (.41 and .34, respectively), whereas the correlation between IQ Estimate and appreciation (.14), although highly significant because of the large n involved, is meaningfully absent ($r^2 = .0196$). This may reflect the weaker reliability of the appreciation measure because of the less defined scoring guidelines as noted in the Manual, the independence between appreciation and conventional IQ, the premorbid nature of the IQ estimate (i.e., furnished by Information and Picture Completion as opposed to Block Design and Similarities), the unique fluid intelligence demands of this competency component due to task novelty, or some combination of the above.

Additionally, correlation data were reported between the three CST domain scores (for understanding, reasoning, and appreciation) and the MMPI-2 Psychoticism Scale, five Brief Psychiatric Rating Scale-Anchored (Overall & Gorham, 1962), and clinicians’ ratings of competence on 283 HI participants. Understanding was most strongly and negatively associated with Brief Psychiatric Rating Scale-Anchored Psychoticism ($r = –.40$) and significantly related to all other scales except Brief Psychiatric Rating Scale-Anchored hostility ($r = –.01$, NS). Counterintuitively, depression ratings evidenced modest but significant (.18) positive correlations with all three CST domains. This may reflect the essential dichotomous nature of participants in the study, that is, those suffering from psychotic (impaired) versus depressive (relatively intact) symptoms and diagnoses. The psychometric measures of psychoticism were more disruptive to fitness than was depression. In summary, the MacCAT-CA appears to have convergent validity with independent scales reflecting clinical states that would theoretically be associated with competency status and ratings made by clinicians.

Finally, the Manual presents predictive validity of the MacCAT-CA in terms of its classification utility (their Table 13). This is accomplished by means of two sequential logistical regression analyses, the first of which compared both jail groups (JU and JT) who are presumed competent against the HI presumed incompetent group. The three steps consisted of demographic/historical variables, clinical status predictors, and lastly, the MacCAT-CA measures. Classification of the “presumed competent” was only marginally assisted by the MacCAT-CA measures in Step 3 (89.1% correct vs. 86.6% in Step 2). Identification of the “presumed incompetent” went from 59.5% to 63.8% between Steps 2 and 3. This appears to argue that addition of the MacCAT-CA variables lend modestly to the demographic/historical and clinical status predictors typically accessed by forensic clinicians. These numbers also imply that about 11% of referrals would be misclassified as fit while they were actually unfit and that approximately 36% of the fit would not be screened out by use of the instrument. Total classification accuracy rose marginally from 76.0% to 79.3% with addition of the MacCAT-CA measures.

A second sequential logistical regression analysis utilized only the JT group (i.e., receiving mental health services in jail) as opposed to including the JU (unscreened jail) group to be differentiated from the HI group. In this second analysis, the MacCAT-CA variables did considerably better in increasing accuracy from 66.7% in Step 2 to 77% in Step 3 for those presumed competent but marginally lowered the classification accuracy from 71.7% in Step 2 to 69.9% in Step 3 for the presumed incompetent. The total classification accuracy then rose...
from 69.4% in Step 2 to 73.2% in Step 3. What this means is that when it comes to identifying presumed incompetent individuals, the MacCAT-CA is marginally less effective than combined consideration of demographic/historical predictors and clinical status predictors, the state of affairs that the rigorous test of CST was supposed to afford.

Additional Research on the MacCAT-FP. Akinkunmi (2002) evaluated the MacCAT-FP using two groups of prisoners, a group with mental illness and one without. Akinkunmi’s purpose was to determine the practicality of administering the MacCAT-FP to groups of prisoners and to evaluate the instruments internal validity and interrater reliability. He also sought to resolve whether the instrument could distinguish fit from unfit individuals, as judged by forensic psychiatrists, and whether it was sensitive to changes over time.

Akinkunmi (2002) first demonstrated that the MacCAT-FP was practical to administer whether the prisoners suffered from mental illness or not. The instrument was shown to have good interrater reliability, was internally consistent, and reflected differences in the mental states and competence capacities of prisoners with and without mental illness. Furthermore, the MacCAT-FP appeared to be sensitive to changes in these mental states and capacities over time. Akinkunmi concluded that these data support the notion that the MacCAT-FP is a step in the right direction to creating an objective measure of fitness.

Recent Research Findings Comparing the FIT-R and MacCAT-CA. Zapf and Roesch (2001) have published a study comparing decisions arrived at through the use of the MacCAT-CA and the FIT-R. Zapf and Roesch reiterated the discrepancies between the broader US and narrower Canadian competency standards noted above and utilized these two instruments as reflective of the two standards to determine (a) the extent to which the legal standards would result in different decisions empirically and (b) the locus of domain responsible for any noted empirical decision discrepancies. They compared outcomes on the two instruments on 100 competent to consent to participate and consenting referral for fitness evaluations at a mental health facility in British Columbia. They found moderately good agreement between the two instruments (kappa = .513), due primarily to the 48 individuals found unimpaired on any domains of the two instruments. However, they found a greater number of impaired individuals on the MacCAT-CA (48%) than on the FIT-R (32%). There was a small group of four individuals who demonstrated impairment on the FIT-R but were unimpaired on the MacCAT-CA.

This result deserves comment. First, most CST assessments show a clear separation between impaired and unimpaired individuals, so that good concordance rates would be expected. For example, patients in psychotic states are most typically unfit. This is not a difficult determination for an experienced and competent forensic clinician. So although a psychotic state is not a criterion for fitness per se, its presence will heavily affect attentional and recall abilities, and generally precluding a finding of fitness, lead to good interrater reliability regardless of the fitness minutiae being similarly addressed by both instruments. Second, the MacCAT identified exactly 50% more individuals as unfit relative to the FIT-R, despite the relatively intensive inquiry afforded by the FIT-R. This likely reflects the additional reasoning stringencies addressed by the MacCAT. This point will be discussed more fully below.

Critique. The MacCAT represents an innovative approach to evaluating CST. It addresses the more recent reasoning and decision-making capacities that apply in some American jurisdictions (see General Discussion below). It has
generated a reasonable amount of empirical research since its publication that supports its concurrent validity. It has not yet generated research examining construct validity with respect to cognitive and neuropsychological intactness/deficits, but that may come in time.

The Evaluation of Competency to Stand Trial-Revised. The ECST-R was developed as a standardized interview for assessing the underlying dimensions of CST via the Dusky standard. The authors also claim that the ECST-R was also the first competency measure designed to detect whether individuals are feigning incompetence (but see NFQ below). The ECST-R contains 18 items and three competency scales that correspond to the Dusky prongs: Consult-With-Counsel (CWC), Factual Understanding of Courtroom Proceedings, and Rational Understanding of Courtroom Proceedings. This instrument also includes 28 items and five scales for Atypical Presentation (ATP): realistic, psychotic, nonpsychotic, impairments, and both (Psychotic and Nonpsychotic combined).

The ECST-R is composed of four main sections with an additional small set of background questions (Rogers, Grandjean, Tilbrook, Vitacco, & Sewell, 2001). The CWC scale evaluates the nature of the defendant–attorney relationship addressing (a) perceptions of the relationship, (b) defendant’s expectations of the attorney, (c) defendant’s understanding of the attorney’s expectations, (d) resolving disagreements, and (e) special means of communicating with the attorney. Within these boundaries, the clinician rates psychotic interference or other impairments that potentially compromise the defendant’s trust, understanding, and communication with his or her defense counsel (Rogers et al., 2001). The Factual Understanding scale addresses the role of key figures in courtroom proceedings. As many defendants respond with incomplete information, optional probes are provided to elicit missing information. The Rational Understanding scale addresses the capacity to make relatively unimpaired decisions (e.g., communications with prosecutors or accepting a plea bargain) and logical judgments about the case (e.g., worst outcome and most likely outcome) (Rogers et al., 2001). Lastly, the Atypical Presentation scale is composed of 28 items that include a subset of items to evaluate the potential of feigned CST.

Reliability. Rogers et al. (2001) reported ECST reliability data obtained in Tillbrook’s unpublished Master’s and Doctoral theses. The Master’s data were obtained from 25 referrals from a county jail and showed a moderate level of agreement (89.3%; median phi coefficient = .75) with data from his PhD study. Internal reliability obtained from 33 inpatient competency referrals was quite impressive for the overall scale (coefficient alpha = .88) but more so for the Factual Understanding (alpha = .92) than the Rational Understanding (alpha = .75) subscale. The CWC scale had an unacceptably low level of internal reliability (alpha = .33). Taken together, these results suggest that the ECST-R has strong internal reliability as a whole, but the CWC scale may not adequately reflect that singular dimension of CST.

Recent Research on the ECST-R. Rogers et al. (2003) examined the three competency models (discrete abilities, domains, and cognitive complexity) with maximum likelihood confirmatory factor analyses. They used combined data from six forensic and correctional samples. After administering the ECST-R, items representative of the Dusky prongs were used to test the three models. The discreet abilities model was the only one of the three indicating a good fit. This suggests that CST should consider the defendant’s factual understanding of the proceedings, rational understanding of the proceedings, and ability to consult with counsel separately. Based on the confirmatory factor
analysis of Rogers et al., the ECST-T competency scales have commendable alphas (.83 to .89) and interrater reliabilities (.97 to .98).

Rogers, Jackson, Sewell, and Harrison (2004) examined the effectiveness of the ATP scales as screens for feigning incompetence to stand trial, as well as a systematic evaluation of feigned incompetence through a simulation design. Very large effect sizes for feigners as compared to jail controls were discovered for the ATP scales. The authors concluded that the ECST-R ATP scales are homogenous and can be used with confidence as feigning screens in CST evaluations.

**Critique.** The ECST-R has sufficient supportive data to justify its use in applied settings, and this case will be strengthened as more studies with larger clinical samples are published. Its internal consistency is high. It has the advantage of identifying potential malingering cases. It separately addresses the Dusky and the more recent reasoning abilities standards deemed by some jurisdictions to be necessary for CST.

**The METFORS/Nussbaum Fitness Questionnaire.** The METFORS Fitness Questionnaire (NFQ, Nussbaum et al., 1998, and recently renamed the Nussbaum Fitness Questionnaire [NFQ]) is a 19-item self-report measure focusing on legal issues typically addressed during fitness interviews. It is primarily intended as a screening tool to identify potentially unfit individuals from large numbers of potentially unfit/incompetent individuals. Seventeen multiple-choice items exist with six response options for each. Of these options, only a single option is deemed “correct,” three are incorrect, but not unreasonably so, while the remaining two are dubbed “absurd” in light of the patent ridiculousness of their content. The remaining two (non-multiple-choice) items require the respondent to “fill in the blank” asking for current charge(s) and lawyer’s name, two fitness items that preclude a multiple-choice format but are relevant and easily scorable. As an objectively scored written instrument, translations are currently available in Chinese, German, Polish, and French, vitiating the need for scarce and expensive translators.

Constructed as a Permutation Model test (May & Hunter, 1993), binomial probability tables were constructed for the NFQ to evaluate the likelihood of obtaining any number of correct, incorrect, and absurd responses by chance. Using statistical criteria, patients who would be screened out (i.e., not require extensive fitness assessments) are those who score above the statistical cutoff for correct answers and would be responding above chance levels and hypothetically considered fit. Patients performing within chance levels on any of the three-item types (correct, incorrect, and absurd) would require more extensive and specific fitness assessments. Greater than chance endorsement of absurd items raises the possibility of blatant malingering, whereas greater than chance endorsement of incorrect items alerts one to the possibility of subtle malingering (Nussbaum et al., 1998).

The original study (Nussbaum et al., 1998) examined 44 individuals referred by the courts for a brief outpatient assessment in the Brief Assessment Unit (BAU) in a large university-affiliated psychiatric hospital. The individuals were seen by one of three trained graduate student research assistants, were provided with information concerning the research project, and upon agreeing to participate in the study, were administered the MFQ and a brief neuropsychological battery designed to assess three “latent” processing domains reasonably associated with FST, attention/concentration (to monitor and register ongoing court proceedings), cognitive efficiency (to comprehend the basics of the trial), and expressive language production (related to instructing counsel). The Orientation and Mental Control subtests from the Wechsler
Memory Scales tapped attention and concentration. The National Adult Reading Test (NART)-R (as a quick estimate of premorbid intellectual ability), Trail Making Test A (Trails A), the Rey-Osterreith copying task, and the Common Item Estimation Task (Wright & Stuss, 1992) were used to reflect cognitive efficiency, whereas the Controlled Oral Word Association (COWA) Test (a.k.a. the FAS or phonemic fluency) and Sentence Arrangement from the WAIS-R-NI (Kaplan, Fein, Morris, & Delis, 1991) served as crude measures of expressive language ability. The MFQ took less than 5 min to administer and the brief neuropsychological battery less than 20 min. This battery was well tolerated by the patients.

Data were analyzed by classifying all observations into three fitness groups according to psychiatric fitness decisions. This resulted in fit, questionable (The “questionable” group is a legally sanctioned group of individuals who are possibly marginally fit but in the opinion of the psychiatrist, in need of 15–30 days of psychiatric inpatient to achieve reasonable clinical stabilization. While within the confines of psychiatric opinion permitted by Canadian law, this designation is largely humanitarian and pragmatic rather than strictly legal in definition.), and unfit groups with 27, 8, and 9 members, respectively. Across all univariate tests (ANOVAs and three subsequent t-test analyses evaluating comparisons between all of the three different fitness groups), the MFQ number correct was clearly the most powerful variable in group separation. Univariate analyses showed highly significant between-groups differences on all three NFQ scales (correct, incorrect, absurd) and between the fit and unfit groups on Orientation, Mental Control, Trails A, and the NART. From a multivariate perspective, the three groups were highly different (MANOVA, Wilk’s lambda = .217; approximate $F = 2.87$, $df = 24, 60$, $p = .0005$). The Discriminant Analysis resulted in 88.64% correct classification. Seven out of the nine unfit were correctly classified with one classified as unfit and the other as questionable meaning that eight of the nine unfit would have been detected and directed for a more in-depth assessment. Six of eight members of the questionable group were correctly classified, with two misclassified as fit. At least some of this group may actually be fit and included based on the psychiatrist’s humanitarian concerns rather than legal competency. These results would seem to provide reasonable detection of actual unfitness. Twenty-six of 27 fits were correctly classified with only one misclassified as questionable. In summary, if in this cohort of 44 probands, the discriminant function was used for screening and all unfit and questionable group members were referred for more detailed FST assessments, only 1 “true” unfit and only 1 questionable group member would have been missed, 26 unnecessary detailed assessments (representing 60% of the total) would have been avoided, while 14 detailed evaluations that might have been unnecessary would have erroneously evaded the screen. Aside from the 60% immediate monetary savings, the justice system would proceed more rapidly reducing delays and generating secondary savings in reduced detention periods.

Current Studies. Study 1: NFQ performance at Toronto’s Old City Hall, Mental Health Court. Purpose and hypotheses. The purpose of the Old City Hall study was a conceptual replication of the earlier METFORS BAU study with a few improvements in mind. First, it was realized that few areas had the luxury of a hospital-centered BAU with psychometrist resources to undertake this testing. As noted above, many fitness assessments are being conducted in jail cells or court houses. Computerized tests are now available to furnish psychometric data that do not require psychometric training. We also wanted to add a second
semistructured instrument to assess fitness, that would be, along with the NFQ and neuropsychological data, double blind with respect to the psychiatric and judicial decisions. Lastly, we wanted to examine a larger number of individuals.

**Measures.** We used the Mississippi State Hospital Version of the Georgia State Court Competency Screening Test-Canadian Adaptation (GSCCST-CA; Bagby, undated) as a methodologically dissimilar (semistructured interview) CST measure. The Integrated Visual and Auditory (IVA; Sanford & Turner, 1994) was used to measure attentional ability and cognitive impulsivity. The computerized neuropsychological battery MicroCog (Powell, Kaplan, Whitta, Catlin, & Funkenstein, 1993) was also used but proved too difficult for virtually all of the participants. Consequently, MicroCog results were not analyzed, and aside from noting its place in terms of procedures, it will not be discussed further.

**Procedures.** Participants were approached in the individual’s holding cell by the researcher (SM). After relevant information was presented and informed consent obtained, the NFQ and then the GSCCST-CA were administered in the individual’s cell, with test scoring taking place at a later time. Then, the individual was led approximately 80 feet to the office containing a desk and computer where the computerized IVA and MicroCog were administered. The computerized tests were followed by administration of the Shipley Institute for Living scale (Zachary, 1986) as a measure of current verbal ability and the NART-2 as an efficient estimate of premorbid intellectual ability. Upon completion of the NART-2, the participant was returned to his/her cell. The computerized tests took approximately 18 and 25 min each, whereas the Shipley Institute for Living scale required about 12 min and the NART-2 about 2 min. Aside from MicroCog, these tests were performed with reasonable effort by a sufficient number of participants to allow analysis. One hundred and eight participants completed NFQ and GSCCST-CA. Because of a number of other issues (including but not restricted to English language proficiency, patience, psychotic mental state, suspiciousness, and hostility), a restricted set of data are available for the neuropsychological tests. Consequently, we will present the full-set fitness data first and then related neuropsychological associations with the fitness scores.

**Fitness results.** Of the 143 participants, 119 were found fit by the attending psychiatrists, 9 fell into the questionable category, and 15 deemed unfit. Table 1 provides raw scores on the NFQ scales and the GSCCST-CA, showing clearly that both these instruments provide very different values to differentiate the fit and unfit groups. The questionable group mean is much closer to the fit group, reinforcing the notion that many members of this group are likely fit in a strictly legal sense but presumably benefited from the afforded treatment. It is also noteworthy that the standard deviation values for the NFQ are relatively homogeneous across the fitness groups, despite the differences in numbers. The GSCCST-CA appeared to provide somewhat better separation between the questionable and the fit groups than the NFQ.

The statistical evaluations associated with these raw score differences are provided in Table 2. All are highly significant and demonstrate the concurrent validity of the NFQ and the GSCCST-CA.

Correlations among the various NFQ scales, the GCCST-CA, and psychiatric decisions (scored 1, 2, and 3 for unfit, questionable, and fit, respectively) show strong positive associations between the NFQ number correct and psychiatric decisions (.722) and a marginally smaller observed association between the GCCST-CA and the psychiatrists’ decisions.
The negative sign between the NFQ incorrect/absurd and the three direct fitness indicators (NFQ correct, GCCST-CA, and psychiatric decisions) follows their reflections of "unfitness" rather than fitness. Thus, the NFQ demonstrates good concurrent validity with both a semistructured interview (across method and some content variance) and independent psychiatric opinion. It is also interesting to note that despite the method variance and the potentially cognitively and emotionally unstable, alienated, and disinterested individuals in a relatively stressful context, these fitness measures showed remarkably high levels of association ($r = .811$) between the NFQ number correct and the GCCST-CA.

Additional insights regarding the fitness construct and integrity of the patients’ responses can be gleaned from a perusal of the neuropsychological and historical measures included in Table 3. For example, Years of education (YOE) is significantly correlated with the NART, a (perhaps imperfect but still useful) measure of premorbid intelligence. YOE is also significantly related to cognitive impulsivity (IVA Response Control Quotient [IVARCQ]) and to a lesser extent, attentional ability (IVA Attention Quotient [IVAAQ]). Fitness itself, whether assessed by the NFQ, GCCST-CA, or psychiatric opinion is uninfluenced by the historical past as reflected by YOE. Alternatively, fitness is related to present measures of cognitive impulsivity (IVA correct $r \sim .30$, $p = .01$) and manifested more strongly in the relationship with absurd responses ($r = -.377$, $p = .006$). This is reasonable because as impulsive individuals lose patience going down the list, their tendency to stop meaningful processing and select an item at random increases and plays into the 33% probability of selecting an absurd option. This is also evident in the very similar correlation between the GCCST-CA and the IVARCQ ($r = .304$, $p = .009$) and somewhat more strongly with psychiatric decision making ($r = .427$, $p = .000$).

Finally, a contrast exists between this measure of impulsivity and the IVA attentional quotient in that unlike the case with the IVA measure of cognitive impulsivity, the IVAAQ appears completely unrelated to psychiatric fitness decisions ($r = .163$, $p = .171$). However, this actually makes perfect sense if one considers that the psychiatrist will be entirely cognizant of an individual’s tendency toward impulsive acting out, which may be linked to an understood fitness defect, unmanageable behavior. However, in a typical interview setting, the psychiatrist focuses the patient’s attention with continuous direct questions. In effect then, the psychiatrist acts as the attentional

<table>
<thead>
<tr>
<th>Fitness decision of psychiatrist</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NFQ number correct</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfit</td>
<td>7.07</td>
<td>4.41</td>
<td>15</td>
</tr>
<tr>
<td>Questionable</td>
<td>14.56</td>
<td>3.09</td>
<td>9</td>
</tr>
<tr>
<td>Fit</td>
<td>17.27</td>
<td>2.82</td>
<td>118</td>
</tr>
<tr>
<td>Total</td>
<td>16.03</td>
<td>4.37</td>
<td>142</td>
</tr>
<tr>
<td><strong>NFQ number incorrect</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfit</td>
<td>8.73</td>
<td>4.58</td>
<td>15</td>
</tr>
<tr>
<td>Questionable</td>
<td>3.67</td>
<td>4.84</td>
<td>9</td>
</tr>
<tr>
<td>Fit</td>
<td>1.54</td>
<td>2.71</td>
<td>118</td>
</tr>
<tr>
<td>Total</td>
<td>2.60</td>
<td>3.81</td>
<td>142</td>
</tr>
<tr>
<td><strong>NFQ number absurd</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfit</td>
<td>2.93</td>
<td>3.80</td>
<td>15</td>
</tr>
<tr>
<td>Questionable</td>
<td>.67</td>
<td>1.11</td>
<td>9</td>
</tr>
<tr>
<td>Fit</td>
<td>.19</td>
<td>0.66</td>
<td>118</td>
</tr>
<tr>
<td>Total</td>
<td>.51</td>
<td>1.60</td>
<td>142</td>
</tr>
<tr>
<td><strong>Georgia</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfit</td>
<td>15.20</td>
<td>7.81</td>
<td>15</td>
</tr>
<tr>
<td>Questionable</td>
<td>25.67</td>
<td>8.97</td>
<td>9</td>
</tr>
<tr>
<td>Fit</td>
<td>35.09</td>
<td>6.07</td>
<td>118</td>
</tr>
<tr>
<td>Total</td>
<td>32.38</td>
<td>9.03</td>
<td>142</td>
</tr>
</tbody>
</table>
navigation system for the patient, obscuring atten-
tional deficits that become evident in a less
structured context. It may be that the research
assistant administering the GSCCST-CA was not
as directive as the psychiatrist because he was
not subject to the same time constraints.

Because this study was purely a research effort,
we did not bring to light cases that should have
been flagged for more detailed evaluations of ma-
lingering by the SIRS (Rogers, Bagby, & Dickens,
1992) or similar instruments. Space does not per-
mit additional tables displaying frequencies of
the numbers of correct, incorrect, and absurd
NFQ choices. However, 17 correct responses
(out of 19) marked the 42.8 cumulative fre-
quency point. Three incorrect responses marked
the 76th cumulative frequency point, and 84%
of the respondents did not select a single absurd
item. Were the NFQ used as a screen, cases pro-
ducing correct/incorrect/absurd choices of 6, 3,
and 10; or 0, 7, and 12; or 3, 13, and 3, respec-
tively, would have been followed up with
appropriate scrutiny.

Summary of Study 1. This study demonstrates
that the NFQ, a self-report scale requiring about
5 min of an individual’s time to complete and
less than 2 min to score provides very similar
data upon which to base opinions regarding
FST as psychiatric opinion or a formal semi-
structured interview in an in situ study. The
NFQ shows expected relationships with neuro-
cognitive measures and is not related to histor-
ical information that should be irrelevant to
FST. The potential of the NFQ to flag cases
for investigation of malingering is yet unex-
loited and will remain so until it is used in
practice rather than as a research tool. Al-
though not advocating that an ultimate FST de-
cision be reached on the basis of this (or any
other) self-report scale, an appropriate use
for the NFQ will be presented and justified
in the General Discussion section below.

Study 2: Concurrent validity and utility of the
NFQ with developmentally delayed individuals.
Assessing the FST in mentally disordered indi-
viduals can be complex but that level of challenge
is dwarfed in assessing FST in the developmen-
tally delayed (DD). DD individuals frequently
possess limitations in attention, instructional
comprehension, and ability to communicate what
they think and do. Because they comprise a min-
imal percentage of the criminal population
(Nussbaum, 2006), it is not surprising that only
one test, the 50-item Competence Assessment
for Standing Trial for Defendants with Mental Re-
tardation (CAST-MR; Everington & Luckasson,
1989), has been designed and validated specifi-
cally for evaluating DD individuals for FST. To
explore the utility of the NFQ to service this
population, this second study provides both pre-
liminary normative data on the NFQ with a DD

<table>
<thead>
<tr>
<th>Source</th>
<th>Dependent variable</th>
<th>Type III sum of square</th>
<th>df</th>
<th>Mean square</th>
<th>F</th>
<th>Significance</th>
<th>Partial eta squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitness decision by the psychiatrist (fit, questionable, unfit)</td>
<td>NFQ number correct</td>
<td>1411.5</td>
<td>2</td>
<td>705.7</td>
<td>76.3</td>
<td>.000</td>
<td>.523</td>
</tr>
<tr>
<td></td>
<td>NFQ number incorrect</td>
<td>710.9</td>
<td>2</td>
<td>355.5</td>
<td>36.8</td>
<td>.000</td>
<td>.407</td>
</tr>
<tr>
<td></td>
<td>NFQ number absurd</td>
<td>100.0</td>
<td>2</td>
<td>50.0</td>
<td>26.4</td>
<td>.000</td>
<td>.276</td>
</tr>
<tr>
<td>Georgia</td>
<td></td>
<td>5691.7</td>
<td>2</td>
<td>2845.4</td>
<td>68.0</td>
<td>.000</td>
<td>.495</td>
</tr>
</tbody>
</table>
### TABLE 3. Correlation Coefficients between the Nussbaum Fitness Questionnaire (NFQ) Scores, the Georgia Court Competency Screening Test (CCST), Canadian Adaptation Scores, the Fitness Decision of the Psychiatrist, Years of Education, the National Adult Reading Test (NART), the Integrated Visual and Auditory Continuous Performance Test (IVA) Full Scale Response Control Quotient (IVARCQ), and the IVA Full Scale Attention Quotient (IVAAQ)

<table>
<thead>
<tr>
<th></th>
<th>NFQ number correct</th>
<th>NFQ number incorrect</th>
<th>NFQ number absurd</th>
<th>Georgia CCST</th>
<th>Fitness decision of the psychiatrist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson correlation</td>
<td>1.00</td>
<td>-.936&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-.623&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.811&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.723&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Significance (two tailed)</td>
<td>145</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>N</td>
<td>145</td>
<td>145</td>
<td>143</td>
<td>145</td>
<td>145</td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>-.936&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.00</td>
<td>.319&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.775&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-.650&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Significance (two tailed)</td>
<td>.000</td>
<td>145</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>N</td>
<td>145</td>
<td>145</td>
<td>143</td>
<td>145</td>
<td>145</td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>-.623&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.319&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.00</td>
<td>-.492&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-.491&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Significance (two tailed)</td>
<td>.000</td>
<td>.000</td>
<td>145</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>N</td>
<td>145</td>
<td>145</td>
<td>143</td>
<td>145</td>
<td>145</td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.811&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-.775&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-.492&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.00</td>
<td>.674&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Significance (two tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>N</td>
<td>143</td>
<td>143</td>
<td>143</td>
<td>146</td>
<td>146</td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.723&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-.650&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-.491&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.674&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.00</td>
</tr>
<tr>
<td>Significance (two tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>149</td>
</tr>
<tr>
<td>N</td>
<td>145</td>
<td>145</td>
<td>145</td>
<td>146</td>
<td></td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.003</td>
<td>.003</td>
<td>-.043</td>
<td>.095</td>
<td>.103</td>
</tr>
<tr>
<td>Significance (two tailed)</td>
<td>.977</td>
<td>.766</td>
<td>.691</td>
<td>.380</td>
<td>.339</td>
</tr>
<tr>
<td>N</td>
<td>86</td>
<td>86</td>
<td>86</td>
<td>88</td>
<td>88</td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.258&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-.196</td>
<td>-.246&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.194</td>
<td>.093</td>
</tr>
<tr>
<td>Significance (two tailed)</td>
<td>.012</td>
<td>.058</td>
<td>.017</td>
<td>.059</td>
<td>.366</td>
</tr>
<tr>
<td>N</td>
<td>94</td>
<td>94</td>
<td>94</td>
<td>95</td>
<td>96</td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.289&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-.205</td>
<td>-.358&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.304&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.427&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Significance (two tailed)</td>
<td>.012</td>
<td>.089</td>
<td>.002</td>
<td>.009</td>
<td>.000</td>
</tr>
<tr>
<td>N</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>72</td>
<td>73</td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.221</td>
<td>-.230</td>
<td>-.139</td>
<td>.312&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.163</td>
</tr>
<tr>
<td>Significance (two tailed)</td>
<td>.069</td>
<td>.058</td>
<td>.255</td>
<td>.008</td>
<td>.171</td>
</tr>
<tr>
<td>N</td>
<td>69</td>
<td>69</td>
<td>69</td>
<td>72</td>
<td>72</td>
</tr>
</tbody>
</table>

<sup>a</sup>Correlation is significant at the .01 level (two tailed).

<sup>b</sup>Correlation is significant at the .05 level (two tailed).
sample and concurrent validity data of the NFQ referenced against the CAST-MR.

One reason that DD individuals may appear unfit, even if they have the requisite knowledge to participate in a relatively simple trial, is a lack of verbal facility to express this knowledge. Smith (1993) found that a significant percentage of defendants with mental retardation did not understand definitions of legal terms that were crucial for a defense and favorable outcome. Sigelman, Winer, and Schoenrock (1982) have argued that a multiple-choice format is superior to an open-ended format because it minimizes the demand on respondents who may not have an extensive enough vocabulary to answer favorably and independently.

Purpose and hypotheses. The primary purpose of this study was to examine the convergent validity of the NFQ and the CAST-MR. The first hypothesis was that CAST-MR and NFQ scores would correlate highly (i.e., \( r = .80 \) or higher). That would provide empirical support for the valid application of the NFQ to DD patient groups. It was further hypothesized that a measure of cognitive efficiency, scanning, and tracking (Trails A) would be significantly associated with both the NFQ and CAST-MR. (The sign of the correlation would be negative due to a higher score on the Trails A being reflective of greater impairment.) Third, it was hypothesized that verbal fluency, as measured by the modified COWA Test (see Measures section), would correlate significantly with NFQ and CAST-MR scores.

Measures. Competency Assessment Screening Test-Mental Retardation.

Typically, FST/CST evaluations involve Axis I disorders involving psychotic states due to ongoing hallucinations, delusions, or severe mood disturbances. However, individuals suffering from intellectual limitations are also critically disadvantaged when facing trial. The CAST-MR (Everington & Luckasson, 1989) is a clinician-administered questionnaire, designed as a screening test of CST for individuals with mental retardation. It consists of 50 questions, 40 of which are multiple choice, each with three options, whereas the remaining 10 are fill-in-the-blank questions, which are relevant only if the respondent is currently before the court.

The CAST-MR focuses on three categories: (a) basic legal concepts, (b) skills to assist defense, and (c) understanding case events. The maximum possible scores allocated to each category are 25, 15, and 10 points, respectively. Due to the nature of the sample in the present study, only the first 40 questions were used as none of the subjects were currently before the courts. The CAST-MR has been validated and cross-validated as a screen and has reported test-retest reliability coefficients of \( r = .89 \) and \( r = .90 \) (Everington, 1990; Everington & Dunn, 1995). Neurocognitive tests used included the simple Trails A, a speeded test requiring connecting 26 numerically sequenced circles in order, and Phonemic Fluency, counting number of distinct, non-proper noun words generated in 60 s. Due to the disadvantaged nature of the sample, only the less complex Trails A and the letter “P” were used in the two tests, respectively. P was chosen as it produces the greatest number of generated responses in the English language (Borkowski, Benton, & Spreen, 1967). A preliminary trial with the letter “S” was used as a warm-up and manipulation check to ensure that the individual understood the task.

Procedures. Participants consisted of individuals identified by a community-based agencies that deal with DD adults (experimental group). Additionally, a community control group was administered the same procedures, but as these are not the focus of this study, those data, available from the first author, are not discussed further. Participants were interviewed
by the primary researcher (IT) who explained the nature of the study as investigating the knowledge of courtroom proceedings. The primary researcher was not involved in any way with the subjects’ case management. Participants were informed that they would answer questions about their knowledge of courtroom proceedings. Participants were further informed they would complete two questionnaires about court (NFQ, CAST-MR), they would draw a line to connect numbers (Trails A), and to say as many words as they could that start with the same sound (COWA). Participants were informed that they would receive either $2.00 or a national (6/49) lottery ticket and that total participation time would be approximately 45 min. Participants were also informed that no individual results would be provided but rather information would be available after the completion of the entire study if requested. No participants withdrew or expressed concern with any of the procedures.

The subjects from both conditions were administered the two competency questionnaires (NFQ and CAST-MR) and the two measures of cognitive efficiencies, the COWA and Trails A. It was anticipated that some experimental participants would not have adequate reading capacity for the measures; consequently, the instruments were read to all participants.

The NFQ was modified in its administration and evaluation. The two fill-in-the-blank questions, plus one multiple-choice question concerning one’s charges, which were not appropriate for this sample, were omitted. This reduced the maximum possible score to 16. The NFQ took approximately 10 min to administer. For the evaluation of decision agreement, a regression equation, predicting NFQ from the CAST-MR, was developed and the NFQ score corresponding to a CAST score of 24 was used as a cutoff.

Administration and evaluation of the CAST-MR was identical to the NFQ. Only the first two categories of the questionnaire were relevant to this sample. The third category “understanding of case events,” was not administered, as subjects were not currently charged, rendering this administration a modified CAST-MR. The CAST-MR was scored as follows: one mark was given for each correct response on the two categories, basic legal concepts and skills to assist defense. Following the authors’ specifications, the cutoff for “competency” was 14 (out of a possible 25) for an understanding of basic legal concepts and 10 (out of a possible 15) for skills to assist defense. Therefore, any participant who scored 24 or more was deemed to have qualified for a finding of competency. The CAST-MR took approximately 15 min to administer.

Design and statistical analyses. Pearson r’s were employed as the correlational statistic because the measures provide ratio scale data (number correct on the fitness measures and time taken to completion on Trails A). Differences between group correlations were evaluated for significance by the r to Z transformation (Hayes, 1973). A secondary focus was on the between-group differences, that is, experimental versus control. A MANOVA was used to evaluate experiment wise Type I errors, followed by univariate ANOVAs for each variable.

Ethical considerations and statement of risk. The present study received ethical approval from the Research Ethics committees of York/Atkinson University, Surrey Place Centre, The Reena Foundation, and York Community Services. Informed consent was obtained from participants by the researcher. As this sample consisted of DD individuals, substitute consent was obtained from each subject as well as his/her legal guardian along with the agency responsible for their care. Participants’ confidentiality was maintained by identifying
research data by research number only. Participants’ intelligence scores were not provided by the agencies because of privacy and confidentiality concerns. As an unfunded undergraduate thesis for Mr Turner, the project did not have the resources to undertake additional psychometric evaluations, although IQ scores would have been a very useful construct to control for.

**Results**

**DD Group Correlations**

The major focus of the paper centered on the group of DD individuals. The calculated correlation between the CAST-MR and the NFQ in this DD group was .86 and was in accordance with the hypothesis that the correlation coefficient of primary interest would exceed .80. The $r^2$ of .74 indicates that approximately 75% of the variance is common to the two measures, arguing for the construct validity of the NFQ and a reflection of fitness knowledge in DD individuals. This robust $r$ value is especially impressive if one considers that the sample represents a truncated range of general intellectual ability and that there is considerable method variance (multiple choice vs. semistructured interview formats) between the two measures.

**Summary of Study**

The Major finding of this study is that a very robust correlation exists between the NFQ and the CAST-MR in the DD group. This demonstrates good concurrent validity of the NFQ and CAST-MR within a population that is generally difficult to assess for fitness. This is especially so in light of the method variance between the two assessment tools noted above. Additionally, good decision agreement, especially with a modified cutoff, led to very good sensitivity, which corroborated the NFQ’s ability as a screen. This study should be considered preliminary and does not alone justify incautious use of the NFQ in FST evaluations. First, the participants were not facing any charges, meaning that a number of items were irrelevant. This is partly responsible for the depressed NFQ and CAST-MR scores relative to what might be expected from a DD sample that has broken the law. The generalizability of these correlational findings remains unknown. The sample size, although reasonable for a pilot study on a small and atypical population, is far from overwhelming. The researcher collected both the CAST-MR and the NFQ interactively and the study was therefore not “blind.” Consequently, although the NFQ might provide a method of eliciting relevant information from DD individuals being assessed for FST, use of a better established instrument should be used to augment clinical opinion at this point in time.

**General Discussion**

**The Fitness Construct**

I have practiced forensic psychology for 20 years in a jurisdiction (Ontario, Canada) that subscribes to a minimalist (limited cognitive) standard. There have been instances over the years when I have thought that an individual, while technically fit, was really ill-prepared to participate meaningfully at his trial. However, there is a clear remedy to this in the criterion requiring counsel to receive clear and meaningful instruction from the client. In these instances, the lawyer has typically advised the court that he/she could not receive instruction and an inpatient placement, with a court directive to “treat until fit” (or treat to fit), was ordered. Additionally, the psychiatrist has the option of declaring
the accused of questionable fitness, triggering a 30-day inpatient stay and a complete psychiatric workup. Thus, practical “fail-safe” mechanisms are available for appropriate, but they must be utilized judiciously to work.

It would appear that the extensive fitness criteria espoused following Judge Bonnie’s landmark publication has resulted in “Mission Creep.” Clearly, fundamental justice requires that individuals must possess sufficient ability to provide reasonable direction to their attorney within the confines of their case. Otherwise, they cannot be meaningfully tried as they are unable to demonstrate potential innocence or advocate for a best outcome even if actually guilty. The question is what constitutes a sufficient ability to provide reasonable direction.

From our perspective, what began as an attempt to help the worst cases has continued to evolve well beyond the reasonable. Items requiring knowledge of optimal legal strategies are likely beyond the ken of many non-mentally disordered criminals and even noncriminals. How many non-attorney readers of this article would feel comfortable in plotting their “optimal legal strategies” in an important case? Presumably, this is why attorneys attend law school, pass bar examinations, and when not working for “legal aid” are entitled to reasonably good remuneration for their advice. The individual accused, provided that he/she can meaningfully instruct counsel, should not require reasoning and decision-making capacities approximating those of a trained legal professional.

A corollary might be that such a standard should extend to all defendants, not only those suffering from a mental disorder. There are likely many non-mentally disordered individuals whose intellectual, educational, or cultural limitations render them disadvantaged in trial scenarios and consequently unable to help mount an optimal defense. Indeed, how many accused of murder have the financial ability to hire a “dream team” of lawyers who might present an optimal defense? Although this is obviously not an issue that will be resolved in the discussion section of an empirically focused paper, this brief discussion is intended more to stimulate thought and debate rather than draw a definitive conclusion.

Unstructured Clinical versus Semistructured and Questionnaire Approaches

Those favoring empirically supported structured/semistructured approaches to FST assessments are still subject to a conceptual dilemma concerning that lack of a “gold standard” and ultimate resorting of criterion validity to the opinions of either an FMHP or the judiciary. Judicial opinions are implicitly contaminated by either a clinical or a structured report, rendering them questionable as a pure criterion measure. Instrument developers are struck by a similar conundrum in that their scales are validated against a psychiatric or a judicial decision or their further contaminated interaction. This underscores the need to show some relationship between a fitness decision and some objective measures of abilities clearly within the fitness domain. Neither the purely clinical nor the semi/structured approach has conclusively shown superiority in this regard.

Consequently, the most demanding and heuristic approach may be to combine approaches in the more contentious or uncertain cases. Agreement between opinions derived from divergent methods minimizes the likelihood that a gross error has been committed. A formalized semi-structured approach for cases where the individual is floridly psychotic, severely brain damaged, or entirely noncommunicative is likely unnecessary, needlessly intrusive, and onerous to the accused.
Optimal Strategies for Semistructured and Questionnaire Utilization

Given the above legal and empirical considerations, what might be an optimal strategy for combined utilization of clinical and semi/structured approaches? This hinges on the scope of the system and nature of the individual case. It makes little sense to provide a large number of expensive assessments to minimally impaired individuals by subjecting them to FMHP administered FST examinations or lengthy and extremely detailed semi/structured interviews. However, a brief paper and pencil test with demonstrated validity could be utilized efficiently as a screen. Those cases not passing the screen could be subjected to either a semi/structured or an FMHP FST evaluation, depending on the resources available to the particular system or organization. Finally, for cases not screened out and/or examined initially by an FMHP presenting some hesitation, a second investigation involving any of the above noted instruments would be suggested. Convergent information would buttress the FMHP opinion, and divergent information would initiate a very careful set of considerations and investigations before arriving at a definitive conclusion. Perhaps that answer might be “wrong,” but because the fitness construct is a social construction, perhaps the best we can expect is that careful and thoughtful consideration be given to this decision that can have extremely important consequences for the lives of a vulnerable population.

Suggestions for Practice

In light of the above, we suggest that initial screening should proceed on a relatively broad number of individuals with an economical and cost effective screening instrument such as the NFQ. This will save the system needless and costly evaluations and identify potential malingerers that could be assessed with more specific instruments to avoid the serious consequences of false positives. Those screened as unfit could then be interviewed by a forensically trained mental health professional. Individuals clearly below the competency standard could be declared incompetent/unfit, whereas those presenting any complexity for the clinician could then be administered any of the three instruments reviewed above. In this way, a broader number of individuals than presently formally assessed could be screened in the hope of detecting the quietly psychotic inhabitants of detention facilities. Although some policy makers and administrators might initially feel that this would result in undue and unsustainable costs with the detection of yet more mentally disordered offenders, early detection and intervention would provide sustained savings by avoiding recurring criminal behaviors.

Conclusions

FST/CST is a psycholegal construct that is not ephemeral but can be evaluated in an objective and reliable fashion. The construct can be measured by different instruments with good agreement and scores on appropriate measures that correlate with neuropsychological measures providing construct validity to the domain. Structured instruments for CST/FST may be reserved for more difficult evaluations, and it may be efficiently and effectively screened by use of the NFQ. Forensic mental health professionals are appropriately trained to conduct these evaluations, and there is no need to be defensive or reticent in doing them provided one has an appreciation of the legal, clinical, procedural, and psychometric complexities involved within this important process.
Acknowledgments

We wish to acknowledge the generous grant from the Human Services and Justice Coordination Project, Ministry of Health and Long Term Care, Government of Ontario, to the first author that enabled the research at the Mental Health Court, Old City Hall, Toronto, reported in the main study. Conflict of Interest: None declared.

References


Babylonian Talmud, Tractate Moed Katan, 3b.


Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured interview of reported symptoms (SIRS) and professional manual*. Odessa, FL: Psychological Assessment resources.


