Evaluating Competency to Stand Trial with Evidence-Based Practice

Richard Rogers, PhD, and Jill Johansson-Love, PhD

Evaluations for competency to stand trial are distinguished from other areas of forensic consultation by their long history of standardized assessment beginning in the 1970s. As part of a special issue of the *Journal* on evidence-based forensic practice, this article examines three published competency measures: the MacArthur Competence Assessment Tool-Criminal Adjudication (MacCAT-CA), the Evaluation of Competency to Stand Trial-Revised (ECST-R), and the Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR). Using the *Daubert* guidelines as a framework, we examined each competency measure regarding its relevance to the *Dusky* standard and its error and classification rates. The article acknowledges the past polarization of forensic practitioners on acceptance versus rejection of competency measures. It argues that no valuable information, be it clinical acumen or standardized data, should be systematically ignored. Consistent with the American Academy of Psychiatry and the Law Practice Guideline, it recommends the integration of competency interview findings with other sources of data in rendering evidence-based competency determinations.

Evidence-based practice for evaluation of competency to stand trial cannot be considered without first providing a clinical context and legal framework. Clinically, the movement toward empirically based assessments has created important advances, some limitations, and substantial resistance. The *Daubert* standard provides a legal framework for evidence-based practice in the forensic arena. This article begins with an overview of evidence-based practice and the *Daubert* standard, which sets the stage for an extensive examination of competency to stand trial via three competency measures.

Paris[1] ably documents the evolution of psychiatric practice from idiosyncratic clinical inferences and basic research studies to systematic investigations of evidence-based practice. Applied mostly to treatment and treatment outcomes, evidence-based practice is an attempt to evaluate treatment efficacies systematically via randomized control trials and meta-analyses.[2,3] These efforts to revolutionize mental health practices are not without critics,[4,5] who raise problems with research design (e.g., weak outcome measures, diagnostic validity, comorbidity, and subsyndromal cases). Established practitioners sometimes are slighted by evidence-based researchers, who now feel "entitled to criticize and rectify clinical authorities" perhaps motivated by "an iconoclastic or even patricidal tendency" (Ref. 5, p 327). While the phrase "patricidal tendency" is an overreach, it does capture the concerns of seasoned practitioners who see the possibility that their decades of experience will be devalued or even discredited by evidence-based approaches. Moreover, the objectivity of evidence-based researchers has been called into question because they are motivated by payment and publication to produce noteworthy results.[4] The acceptance of evidence-based methods within the psychiatric community is clearly influenced by both concerns regarding research design and polarized professional attitudes. While the bulk of the article addresses research findings, the next two paragraphs outline the equally important topic of professional attitudes.

Professional attitudes are an often overlooked but key component in the acceptance of evidence-based

Dr. Rogers is Professor of Psychology, University of North Texas, Denton, TX. Dr. Johansson-Love is Clinical Assessment Specialist, Dallas County Community Supervision and Corrections, Dallas, TX. Address correspondence to: Richard Rogers, PhD, 1155 Union Circle #311280, Denton, TX 76203-5017. E-mail: rogersr@unt.edu. Dr. Rogers is the principal author of the Evaluation of Competency to Stand Trial-Revised (ECST-R) and receives a royalty for each ECST-R record form and summary sheet administered.

practice. Slade and his colleagues[6] carefully evaluated the acceptance of an empirically based assessment model involving a constellation of standardized measures. Objections by practitioners to using the assessment model have included concerns about its cost (35%), usefulness (38%), duplicated effort (23%), and duration (10%). As evidence of polarized views, three of these same objections were seen by other practitioners as benefits including usefulness (45%), nonduplication of services (25%), and brevity (25%). Lessons from Slade *et al.* can clearly be applied to forensic practice regarding important determinants for the acceptance of evidence-based practice.

Aarons *et al.*[7,8] have gone a step further in studying how professional attitudes toward evidence-based practice are reflected in effective interventions. Although they focused on treatment, several findings may be applicable to forensic practice. The two most salient objections to evidence-based practice were that clinical experience is better than standardized methods and that practitioners know better than researchers. We revisit these objections later in the context of evidence-based competency measures. The next section addresses the admissibility of expert evidence in light of the *Daubert*[9] standard.

Application of the Daubert *Standard*

The Supreme Court, in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,[9] applied scientific principles to the admissibility of scientific evidence. It explicitly rejected the test established in *Frye v. United States*,[10] which relied solely on general acceptance. While serving as gatekeepers, trial judges are to consider the following guidelines under *Daubert*:

1. Ordinarily, a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested.

2. Another pertinent consideration is whether the theory or technique has been subjected to peer review and publication.

3. Additionally, in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error.

4. Finally, "general acceptance" can yet have a bearing on the inquiry. A "reliability assessment does not require, although it does permit, explicit identification of a relevant scientific community and an express determination of a particular degree of acceptance within that community [Ref. 9, pp 593–4].

Guidelines 1 and 3 specifically address scientific methods. Guideline 1 relies on the construct of falsifiability set forth by Popper.[11] Simply put, a conclusion cannot be accepted as true if there is no way that its truth or falsity can be proven—if it has never been tested. With reference to forensic concerns, can the concept be empirically tested and does the research have the potential to disprove the conclusion? Whereas Guideline 1 is more theoretical, Guideline 3 is solidly methodological. Its error rate focuses specifically on the accuracy of measurement, which is affected by reliability and validity.

Daubert and two subsequent Supreme Court cases (*General Electric Co. v. Joiner*[12] and *Kumho Tire Co. v. Carmichael*[13]) are referred to as the *Daubert* trilogy. In *Joiner*, the Court specified that the trial judge would be the arbiter of scientific admissibility and could be overruled based only on the abuse-of-discretion standard. For mental health experts, the practical effect of this ruling is that different trial judges within the same jurisdiction may legitimately reach opposite conclusions about the admissibility of specific methods, such as competency measures.[14] In *Kumho*, the Supreme Court applied the *Daubert* guidelines beyond scientific evidence to all expert testimony. The practical effect of this decision was to prevent experts from circumventing *Daubert* by claiming that their expertise (e.g., clinical practice) was nonscientific. The Court reaffirmed the flexibility in applying the *Daubert* guidelines, which may or may not be relevant in determining the reliability of the expert testimony in a particular case. Welch[15] extensively describes "*Daubert*'s legacy of confusion" in allowing trial judges to apply any or all of the *Daubert* guidelines when admitting expert testimony.

A comprehensive review of the *Daubert* decision is far beyond the scope of this article, given the hundreds of scholarly works in the psychological, medical, and legal literatures. Readers may wish to refer to the Federal Judicial Center[16] and special issues of *Psychology, Public Policy, and Law* (vol. 8, issues 2–4) and the *American Journal of Public Health* (vol. 95, suppl. 1) for a more thorough introduction. For our purposes, we selectively review articles that provide key insights in *Daubert* and examine several examples of how *Daubert* has been applied to standardized measures and legal standards.

Gatowski and her colleagues,[17] in a national study of 400 state trial court judges, found that most judges

(i.e., ranging from 88% to 93%) believed that the individual *Daubert* guidelines were useful in deciding the admissibility of scientific evidence. Not surprisingly, they had the most difficulty in fully understanding those directly involved in scientific method (Guidelines 1 and 3). In contrast, Guidelines 2 and 4 were relatively easy to grasp. Based on her work, we should anticipate that more scientific guidelines will generate greater discrepancies among trial courts.

Researchers and scholars have critically evaluated whether general psychological tests meet the *Daubert* guidelines for admissibility. For example, controversy and debate surround the sufficiency of the Rorschach[18,19] and MCMI-III[20,21] when evaluated according to *Daubert* guidelines. Regarding the MCMI-III, Rogers and his colleagues[22] questioned the admissibility of any measure when the error rate substantially exceeded its accuracy. *Daubert* reviews have also considered several forensic measures for which the adequacy of their psychometric properties has been debated: competency to confess measures[23,24] and the Mental State at the Time of the Offense scale.[25,26]

Within the context of family law, Kelly and Ramsey[27] provide a masterful analysis of validity as it applies to psycholegal constructs and measures, along with a detailed list of specific benchmarks. Researchers and practitioners are likely to find this a valuable resource in evaluating forensic measures.

Author Disclosure

The opening paragraph of this article noted the professional schisms between traditional practice and the growing movement toward evidence-based practice. Among the broad array of criticisms, researchers have been singled out as motivated by personal and professional gain.[5] An alternative view is that traditionalists are equally motivated to avert criticisms of their current clinical practices by researchers. Be that as it may, a brief disclosure from the first author is in order. Rogers has pioneered the use of empirically validated forensic measures for more than two decades, beginning in 1984 with the publication of the R-CRAS (Rogers Criminal Responsibility Assessment Scales)[28] for assessing criminal responsibility and later the Structured Interview of Reported Symptoms (SIRS)[29] for feigned mental disorders. Of particular relevance to this article, he is the principal author of the Evaluation of Competency to Stand Trial-Revised (ECST-R)[30] and receives a royalty of approximately 30 cents for each ECST-R record form and summary sheet administered. Readers can independently evaluate the following analyses of competency measures in light of this disclosure.

Competency to Stand Trial

The standard for competency to stand trial was established by the Supreme Court's decision in *Dusky v. United States*[31] with a one-sentence formulation requiring that the defendant "has sufficient present ability to consult with his lawyer with a reasonable degree of rational understanding—and whether he has a rational as well as factual understanding of the proceedings against him." Rogers and Shuman[14] provide a legal summary of *Dusky*'s three prongs: a rational ability to consult one's own attorney, a factual understanding of the proceedings, and a rational understanding of the proceedings. Practitioners should be familiar with the *Dusky* standard and relevant appellate cases.

Competency to stand trial is especially important to evidence-based forensic practice because of its prevalence; it represents the most common pretrial focal point within the criminal domain of forensic psychiatry. Conservative estimates suggest there are 60,000 competency cases per year, with rates of incompetency often falling in the 20- to 30-percent range.[32] When extrapolated from the number of actively psychotic and mentally disordered inmates,[33] the potential number of competency evaluations could easily be twice this estimate.

Competency evaluations are also relevant to evidence-based forensic practice because of their long history of empirical validation. In his seminal work, Robey[34] proposed in 1965 a standardized checklist for operationalizing competency to stand trial. With NIMH support, Lipsitt and his colleagues[35] developed in 1971 the first standardized competency measure, the Competency Screening Test (CST). It was followed in 1973 by the Competency Assessment Instrument (CAI), developed and validated by McGarry and his team[36] at Harvard Medical School's Laboratory of Community Psychiatry. This historical perspective provides an essential insight: the foundation for evidence-based forensic practice was established while the American Academy of Psychiatry and the Law (AAPL) and its counterpart, the American Academy of Forensic Psychologists, were

still in their infancies. Unlike other forensic concerns, competency to stand trial has been the vanguard of evidence-based practice, championed for decades by prominent forensic psychiatrists and psychologists.

The importance of competency evaluations was recently underscored by the 2007 publication of the AAPL Practice Guideline.[37] This guideline provides a thorough introduction to the legal framework and conceptual basis for conducting these evaluations. While it does not grapple directly with evidence-based practices, the guideline attempts to standardize competency evaluations by recommending 15 specific areas of inquiry. Without providing standardized questions, it provides a nuanced statement that "Assessing and documenting a defendant's functioning usually requires asking specific questions that systematically explore" competency-related abilities (Ref. 37, p S34). Parenthetically, the qualifying term "usually" seems difficult to understand. Nonetheless, the AAPL Task Force recommends the use of specific questions and a systematic examination covering 15 areas of inquiry. Could each forensic psychiatrist or psychologist develop his or her own specific questions and systematic examination of competency? Although theoretically possible, an affirmative response would suggest marked optimism that does not take into account the need to establish the reliability and accuracy of their systematic examinations. A more sound approach would be the integration of clinical interviews with standardized measures. In fact, this approach is embraced by the AAPL Task Force in its summary statement about competency measures: "Instead, psychiatrists should interpret results of testing in light of all other data obtained from clinical interviews and collateral sources" (Ref. 37, p S43).

Evidence-based practice cannot be achieved without standardization. For assessments, the use of reliable and valid measures is the most direct and empirically defensible method of achieving this standardization. The remainder of this article assumes that practitioners will integrate case-specific (clinical interview and collateral information) with nomothetic (standardized results) data. The standardized results, while only one component of competency evaluations, achieve four major objectives by systematizing the evaluation of key points, reducing the subjectivity in recording competency-related information, providing normative comparisons, and

demonstrating the inter-rater reliability of observations and findings. Despite these important contributions to competency assessments, the caution of the AAPL Task Force is well founded; conclusions should not be based only on this source but should reflect a careful integration of multiple sources of data.

Overview of Competency Measures

The first-generation of competency measures was introduced in the 1970s. Of mostly historical interest, first-generation measures have limited data on their psychometric properties, a lack of normative data, and poor correspondence to the relevant legal standard.[38] Although reviews of these measures are readily available,[39] this article focuses more selectively on three published competency measures. Two measures are intended for general competency evaluations: the MacArthur Competence Assessment Tool-Criminal Adjudication (MacCAT-CA)[40] and the ECST-R.[30] The third measure, the Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR),[41] concentrates on defendants with mental retardation. The purpose of these competency measures is to provide standardized data to assist practitioners in reaching empirically based conclusions about elements of competency to stand trial. As noted by one reviewer, it would be utterly naïve to attempt to equate any test or laboratory findings with an ultimate or penultimate legal opinion.

The following subsections provide a brief description of the measures and their development. They are followed by a more in-depth examination of competency measures as a form of evidence-based practice.

MacCAT-CA Description

The MacCAT-CA was not originally developed as a measure of competency to stand trial. Instead, the original MacArthur research was intended to assess a much broader construct of decisional competence via a lengthy research measure, the MacArthur Structured Assessment of the Competencies of Criminal Defendants.[42] It was subsequently shortened and retrofitted for the evaluation of competency to stand trial.

The MacCAT-CA is composed of 22 items that are organized into three scales: understanding (8 items), reasoning (8 items), and appreciation (6 items). Probably because of its original development as a research measure, 16 of the 22 items do not

address the defendant's case. Rather, the Mac-CAT-CA asks the examinee to consider a hypothetical case about two men (Fred and Reggie) and their involvement in a serious, almost deadly, assault following an altercation while playing pool.

The MacCAT-CA has excellent normative data for 446 jail detainees, 249 of whom were receiving mental health services. They were compared with 283 incompetent defendants in a competence restoration program. These normative data were used for clinical interpretation of information from the jail detainees to establish three categories. Minimal or no impairment had assessed deficits that fell within 1 standard deviation (SD) of the presumably competent detainees. Mild impairment was designated as the narrow band of deficits falling between 1 and 1.5 SD. Clinically significant impairment was designated as deficits at and above 1.5 SD. Unfortunately, this approach was unsuccessful for the appreciation scale; the authors simply assigned cut scores to the three categories, based on their own hypotheses regarding delusional thinking.

ECST-R Description

The ECST-R is composed of both competency and feigning scales. Its competency scales parallel the *Dusky* prongs: Consult With Counsel (CWC; six items), Factual Understanding of the Courtroom Proceedings (FAC; six items), and Rational Understanding of the Courtroom Proceedings (RAC; seven items). For feigning, the ECST-R uses Atypical Presentation (ATP) scales that are organized by content (i.e., ATP-Psychotic and ATP-Nonpsychotic) and purported impairment (i.e., ATP-Impairment). Most competency items are scored on five-point ratings: 0, not observed; 1, questionable clinical significance; 2, mild impairment unrelated to competency; 3, moderate impairment that will affect but not by itself impair competency; and 4, severe impairment that substantially impairs competency.

The ECST-R was developed specifically for the purpose of evaluating the *Dusky* prongs. The key components for each prong were assessed by five competency experts via prototypical analysis. Those components retained an average of 6.10 on a 7.00 rating scale of their representativeness. Individual items for the competency scales were developed and pilot tested. The feigning scales were developed by using two primary detection strategies: rare symptoms and symptom severity.

The ECST-R has excellent normative data based on 200 competency referrals and 128 jail detainees. In addition, data were available for comparison purposes for 71 feigners as classified by simulation research or results on the SIRS.[29] Cut scores were developed on the basis of linear T scores, which facilitates their interpretation. One limitation of the ECST-R is that its cut scores have not been validated for defendants with IQs of less than 60. Unlike the MacCAT-CA, which restricts its normative data to presumably competent participants, the ECST-R includes both competent and incompetent defendants in its normative group, thereby mirroring the entire population that it is intended to evaluate. This observation is a likely explanation for the differences in cut scores between the two measures. The ECST-R uses the following classification: 60 to 69 T, moderate impairment, usually associated with competent defendants; 70 to 79 T, severe impairment, which can reflect competent or incompetent defendants; 80 to 89 T, extreme impairment, usually associated with incompetent defendants; and 90 to 110 T, very extreme impairment, almost always associated with incompetent defendants.

CAST-MR Description

The CAST-MR is composed of three competency scales: Basic Legal Concepts (25 multiple-choice questions), Skills to Assist Defense (15 multiple-choice questions), and Understanding Case Events (10 open-ended questions). Basic Legal Concepts is the one most closely aligned with *Dusky*'s factual understanding, whereas skills to assist defense uses hypothetical examples to evaluate the consult-with-counsel prong. Understanding case events asks for detailed recall (e.g., date and witnesses) of the alleged crime and the current criminal charges. Although not a perfect match, this last scale is most closely aligned with factual understanding.

The CAST-MR is an outgrowth of a doctoral dissertation. A small group of 10 professionals (lawyers, administrators, and forensic psychologists) rated the appropriateness of the CAST-MR content. On a five-point scale, the ratings were somewhat variable, with Skills to Assist in Defense reaching an average score of only 3.03 regarding the appropriateness of its content (Ref. 41, p 31).

The CAST-MR is administered as an interview, although examinees are given a copy of the items to facilitate comprehension. According to its authors,

the CAST-MR has a reading level of fourth grade or less, which was calculated by taking two samples, each less than 400 words, and subjecting them to reading estimates.

Descriptive but not normative data are presented from two studies of criminal defendants. A total of 128 criminal defendants compose the following groups: no mental retardation or mental disorder ($n = 46$), mental retardation but no competency evaluation, ($n = 24$), mental retardation and competent ($n = 27$), and mental retardation and incompetent ($n = 31$). The second validation study indicated a moderate agreement (71%) between cut scores and examiner judgment.

Competency Measures and Evidence-Based Practices

With *Daubert* used as the conceptual framework, this section examines competency measures as evidence-based practice. We begin with an evaluation on the congruence between the competency measures and the *Dusky* standard. Next, we examine these measures in light of error and classification rates.

Relevance of Competency Measures

The Supreme Court held in *Daubert* that expert testimony must be relevant to the matter at hand. Citing Federal Rule of Evidence 702, it "requires a valid scientific connection to the pertinent inquiry as a precondition to admissibility" (Ref. 9, p 592). It describes relevance as a matter of "fit"; scientific validity is not sufficient unless it fits the specific matter under consideration by the trial court. For competency determinations, the Supreme Court in *Dusky* established the three prongs for which the "fit" or congruence of scientific evidence must be considered.

Specific factual aspects of cases must also be considered. For example, the three competency measures differ in the extent to which they have been evaluated for pretrial defendants with mental retardation. For scientific validity to be relevant, it must be "sufficiently tied to the facts of the case" (Ref. 9, p 591). Therefore, the following analysis examines the construct validity of competency measures in light of their specific applications to defendant categories.

Table 1 provides a summary of the specific scales on competency measures with descriptive data regarding their type of inquiry and the complexity of their questions. Inquiries can be either case-specific (i.e., the content focuses on the defendant's case) or hypothetical (i.e., the content is unrelated to the defendant's case). Obviously, case-specific data are likely to meet the *Daubert* guideline of being "sufficiently tied to the facts of the case." In contrast, hypothetical data must be examined closely to determine its relevance or fit to a particular defendant's case. For instance, what would be the similarities in MacCAT-CA's aggravated assault between friends and delusionally motivated crimes?

With respect to relevance and fit, three competency measures have the most in common in their assessment of *Dusky*'s factual understanding of the courtroom proceedings. Each evaluates the defendant's understanding of the courtroom personnel and their respective roles at trial. The CAST-MR provides the broadest appraisal of factual understanding with inquiries about common legal terms and basic information regarding verdicts and sentencing. The CAST-MR also has a specific scale for considering the defendant's memory of the offense and subsequent arrest. Recall of these events is likely to be helpful in competency cases in which amnesia plays a central role. The MacCAT-CA also assesses courtroom personnel and then uses a hypothetical case to evaluate criminal charges related to assault and matters such as plea bargaining. Although considered to be factual understanding,[40] this scale also requires rational abilities in deciding on the alternatives. Neither the CAST-MR nor MacCAT-CA assesses defendants' knowledge of their own criminal charges and the severity of these charges. The ECST-R focuses on both courtroom proceedings and defendants' understanding of their own criminal charges.

Forensic practitioners should decide which is most relevant to a particular competency evaluation. As a simple reminder, the CAST-MR has been validated only in defendants with mental retardation; it should not be used for mentally disordered defendants, with or without mental retardation. One strength of the ECST-R is that it both prompts and educates defendants with insufficient responses on factual understanding.

The competency measures are markedly divergent in their assessment of *Dusky*'s consult-with-counsel prong. The MacCAT-CA uses a hypothetical assault to evaluate the defendant's ability to distinguish relevant and irrelevant information and consider

Table 1 Description and Congruence ("Fit") between *Dusky*'s Prongs and Selected Competency Measures

Measure	Scale (*n*)	Type	*M* Length	*Dusky* Prong	Representative Content
MacCAT-CA	Understanding (8)	Hypoth	45.3	Factual understanding	Roles of courtroom personnel; understanding different criminal charges related to assault; understanding convictions and plea bargaining
	Reasoning (8)	Hypoth	39.9	Consult counsel	Distinguishing helpful from unhelpful information to provide to counsel; understanding mitigating factors (e.g., provocation and intoxication); making decisions about plea bargaining
	Appreciation (6)	Case	22.7	Rational understanding	Beliefs about outcome (likelihood of conviction and punishment); perception of lawyer (helpfulness and trust with all information); beliefs about the legal system (fairness and viability of plea bargaining)
ECST-R	FAC: factual	Case	7.2	Rational understanding	Roles of courtroom personnel; understanding the criminal charges
	CWC: counsel	Case	7.7	Consult counsel	Perceptions and expectations of counsel; identifying and resolving disagreements with counsel; impaired communication with counsel
	RAC: rational	Case	8.1		Decision-making about trial; appraisal of different outcomes; problematic courtroom experiences
CAST-MR	Legal concepts (25)	Case	20.3	Factual understanding	The duties of legal professionals in court; common legal terms; specific terms related to sentencing
	Understand case (10)	Case	7.4	Factual understanding	Recall of the crime; recall of the arrest; description of criminal charges
	Assist defense (15)	Both	46.9	Consult counsel	Cooperation with the lawyer; doing what others (e.g. police, inmates, or prosecutors) ask; response to persons (prosecutor or witness) telling lies about the defendant)

M length is the average number of words addressed to the defendant before he is asked to respond; some items include a statement followed by an inquiry. Case is the specific queries about the defendant's case; Hypoth is hypothetical queries unrelated to the defendant; and Both is a combination of case-specific and hypothetical queries.

choices related to matters such as plea bargaining. Therefore, it assesses rational abilities but does not consider the actual defendant-attorney relationship or the ability to communicate rationally. We have found the MacCAT-CA especially useful in competency cases in which the defendant has expressed an interest in serving as his or her own attorney. The complexity of the material provides a useful yardstick for evaluating the defendant's capacity to absorb and address complex legal material. The CAST-MR uses some hypothetical material (e.g., a theft) but mostly relies on material in the defendant's case. It emphasizes the ability of the defendant to cooperate with his counsel, while not acquiescing to others (e.g., police or prosecutors). Although it does not assess the quality of the defendant-attorney relationship directly, it can provide valuable information about the defen-

dant's willingness to cooperate. The ECST-R focuses on the nature of the defendant-attorney relationship; through open-ended questions, it examines the quality of that relationship and the defendant's ability to identify and resolve disagreements in relationship to the trial.

For the rational-understanding prong, both the MacCAT-CA and the ECST-R elicit information about the likely outcome of the case. They differ in that the ECST-R examines how severe psychopathology may affect the defendant's rational abilities. The MacCAT-CA also includes several items about defendants' views and actions toward their attorneys. This information may help with the consult-with-counsel prong. The ECST-R also asks defendants to consider how they might make important decisions about their cases, such as plea bargaining. The focus

Table 2 Reliabilities and Error Rates of the Three Competency Measures

			Competency Measure		
Estimate	Description	Scales	CAST-MR	McCAT-CA	ECST-R
Reliability	Inter-rater reliability establishes the reproducibility of scores when evaluated by two or more experts at the same time; lower estimates equal larger errors.	Competency Feigning	0.90* NA	0.83 NA	0.93 0.996
SEM	Standard error of measurement (SEM) measures the likely variability in the accurate measurement of a single score; larger scores equal greater errors.	Competency Feigning	1.15† NA	1.28 NA	1.28 0.18
95% CI	95% Confidence interval establishes the range of scores possible for a single score that is likely to occur most of the time (i.e., 95 of 100 times); larger scores equal greater errors.	Competency Feigning	2.25 NA	2.51 NA	2.51 0.35

*Inter-rater reliabilities were reported only as percentages; this correlation represents test-retest reliability for nonforensic cases.
†Based on unweighted scale averages ($M = 10.14$, $SD = 3.63$) for four small subsamples of competency cases ($n = 58$).

of the ECST-R inquiries is not on the decision itself but rather on the reasoning underlying the decision.

The foregoing discussion focused on the congruence between competency measures and the *Dusky* standard. Beyond this critically important discussion, the relevance of a measure must also consider its appropriateness for the intended population (i.e., impaired defendants). For example, does the length and complexity of competency questions substantially exceed the defendant's ability to process this information? For normal (unimpaired) persons, the capacity to process information is generally limited to the magic number of 7 ± 2 concepts.[43] For language, individuals use verbal chunking consisting of 6 to 12 syllables per concept.[44] Using the MacCAT-CA as a benchmark with 1.34 syllables per word, the midpoint for unimpaired persons would be: 7 concepts $\times$ 9 syllables $\div$ 1.34 syllables per word = 47.01 words. The lower limit for unimpaired persons is 22.38 words. Defendants with serious mental disorders or mental retardation are likely to have substantial deficits in capacity to process information. In the absence of specific data, one option would be to use the lower limit for normal persons (i.e., ≤ 22 words) as the upper limit for competency measures used with potentially impaired defendants. As summarized in Table 1, two scales of the CAST-MR appear to meet this guideline, with understanding case events being particularly straightforward. In contrast, questions for the assist defense scale include preliminary information that increases the average length to 46.9 words. Likewise, two McCAT-CA scales are also problematic because of their word length: understanding (mean [M] = 45.31

words) and reasoning ($M = 39.88$ words). In direct contrast, the ECST-R took into account word length in the development of its items. As a result, the presented material is typically very short (i.e., fewer than 10 words) on the ECST-R competency scales.

Error Rates and Competency Measures

A major strength of the three competency measures is the excellent data on their reliability and errors in measurement. As summarized in Table 2, trained practitioners are able to achieve a high level of inter-rater reliability on each measure, with exceptional estimates for the CAST-MR ($r = 0.90$) and ECST-R ($r = 0.93$ and 0.996). Because the reliability of traditional interviews cannot be established, the use of these competency measures addresses the scientific reliability of expert evidence.

The *Daubert* guidelines ask that experts address the error rates associated with their methods. One sound approach to ascertaining error rates is to estimate the accuracy of individual scores on competency measures. Calculated as the standard error of measurement (SEM), each competency measure produces small SEMs, indicating a high level of accuracy (Table 2). Especially useful for court reports and subsequent testimony is the 95 percent confidence interval. When an elevated score exceeds the benchmark by the confidence interval, the practitioner can testify regarding a very high likelihood that the defendant meets this classification. As reported in Table 2, expert ratings of defendants that exceed the cut scores by three or more points have at least a 95 percent likelihood of

being accurate. Stated in *Daubert* terms, the error rate is five percent or smaller.

An important consideration in establishing error rates is whether bogus (e.g., malingered) presentations will be mistaken for genuine incompetency. In this regard, the ECST-R is distinguished from the other two competency measures by its highly reliable scales that screen for feigned incompetency. As noted in Table 2, the ECST-R feigning scales have very high reliabilities ($M = 0.996$) and exceptionally small 95 percent confidence levels ($M = 0.35$).

Classifications by Competency Measures

As an outgrowth of the previous section, practitioners must not only consider the relevance of the psycholegal constructs but also the meaning of its classifications. Simply put, how are these classifications established and what is their relevance to the *Dusky* standard? Melton and his colleagues were the first to raise the concern of whether competency measures "appear to permit gross incongruencies between item ratings and scale interpretations" (Ref. 32, p 154). Of interest, that criticism was leveled specifically at the ECST-R rather than being evaluated critically for competency measures in general. We will consider the scale classifications (interpretations) in the subsequent paragraphs.

The CAST-MR test manual provides little guidance for making classification of competent and incompetent defendants with mental retardation. While cautioning that the CAST-MR is only one part of the competence assessment, we note that the mean total score for the defendants with mental retardation was 25.6 for incompetence versus 37.0 for competence. Because of small sample sizes and large variability, they provide the following caution: "only a gross estimate can be made of the degree to which CAST-MR total scores discriminate between groups found to be competent versus those found to be incompetent" (Ref. 41, p 19). In addition, the lack of information about specific prongs is a limiting factor about the CAST-MR classifications.

The MacCAT-CA has the most problems of competency measures in establishing accurate classifications. Obviously, the group of hospitalized legally incompetent defendants should theoretically evidence clinically significant impairment, given their combined psychiatric and legal status. The figures reveal that this is not supported, revealing a flaw in the test. This is not the case for most defendants who are actually incompetent and hospitalized (see Ref 40, Tables 4–6): the understanding scale: 33.2 percent clinically significant impairment, 15.9 percent mild impairment, and 50.9 percent minimal or no impairment; the reasoning scale: 41.3 percent clinically significant impairment, 13.8 percent mild impairment, and 44.9 percent minimal or no impairment; and the appreciation scale: 44.5 percent clinically significant impairment, 9.2 percent mild impairment, and 39.2 percent minimal or no impairment.

Although classifications based on the ECST-R evidence a high concordance with legal outcome (88.9%), classifications by ECST-R scales are based on construct validity and the use of normative data. The ECST-R manual provides extensive data on the accuracy of its measurements. What about the "gross incongruencies" criticism of the ECST-R of Melton and his colleagues[32]? They seem to stem mostly from apparent confusion over the meaning of an ECST-R rating of 3. As previously noted, a rating of 4 shows substantially impaired competency by itself, whereas a rating of 3 shows deficient competency but does not, by itself, show substantially impaired competency. However, the cumulative effects of a 3 rating can indicate substantially impaired competency. Indirectly, the Melton *et al.* commentary did raise a valid question as to whether consistent ratings of 2 (i.e., mild impairment but unrelated to competency) could result in classification as having severe impairment on the ECST-R competency scales. For two scales (FAC and RAC), such ratings would show only moderate impairment, which is typically associated with competent defendants. For the third scale (CWC), it is theoretically possible to score in the severe range based only on ratings of 2. In reviewing the ECST-R normative data, we did not find a single case of any of the competency scales where this occurred. Despite its extreme rarity (i.e., 0 for 356 defendants), practitioners may want to consider quickly screening ECST-R protocols for this remote possibility.

Concluding Remarks

Forensic practitioners should supplement the previous analysis with careful reviews from other researchers and scholars. Grisso[39] provides a thorough review of the CAST-MR and the MacCAT-CA. Although the newest measure, the ECST-R is the only one of these competency measures to be reviewed by

the well-respected Mental Measurements Yearbook.[45,46] By combining these sources, practitioners will become knowledgeable regarding the strengths and limitations of competency measures.

Our informal observations suggest that forensic psychiatrists and psychologists are divided with respect to their use of competency measures. However, the historical divisions between psychiatry and psychology on the use of standardized assessments are gradually disappearing. As evidence of their growing importance, an American Psychiatric Association Task Force undertook a multiyear analysis of psychiatric measures resulting in a comprehensive textbook.[47] Beyond these general trends, specific contributions to competency measures have been multidisciplinary from the early efforts in the 1970s. If not based on disciplines, what accounts for this polarization? We believe that failures of both researchers and practitioners are to blame.

Researchers sometimes overestimate the ability of their standardized measures to evaluate complex clinical constructs. For instance, interview-based competency measures are typically composed of several dozen relevant constructs that are operationally defined. Even with exceptional care, these items can never fully capture the defendant's functioning with respect to the spectrum of competency-related abilities. For example, standardized observations of attorney-client interactions would be valuable. However, efforts in this direction have not been successful. As noted by Melton and his colleagues, "most attorneys have neither the time nor the inclination to observe, much less participate in, competency-to-stand-trial evaluations" (Ref. 32, p 148). Beyond complex content, we suspect there is some professional arrogance arising from the use of sophisticated research designs and psychometric rigor. The "patricidal tendency" of researchers to diminish the contributions of seasoned practitioners may play a relevant role.

Practitioners sometimes exaggerate the limitations of standardized measures while possibly overvaluing their own expertise. Some resistance is encountered from the either-or fallacy wherein practitioners erroneously assume that they must choose between their own individualized methods and psychometrically validated measures. As found by Aarons et al.,[7,8] we suspect there is some professional arrogance arising from views that practitioners are superior to researchers and their standardized methods.

Gutheil and Bursztajn[48] wisely counsel that forensic practitioners avoid even the appearance of "ipse dixitism" with respect to unsubstantiated opinions. Substantiation should embrace an array of relevant sources by knowledgeable experts. As part of this substantiation, reliable and standardized information from competency measures should not be routinely ignored by forensic practitioners. We must tackle directly the professional objections to evidence-based practice. Borrowing from Slade et al.[6]: are these measures useful, nonduplicative, and time-efficient? With professional experience and expertise, practitioners can make informed decisions in selecting the appropriate competency measure to evaluate specific competency-related situations.

References

1. Paris J: Canadian psychiatry across 5 decades: from clinical inference to evidence-based practice. Can J Psychiatry 45:34–9, 2000
2. Dongier M: Evidence-based psychiatry: the pros and cons. Can J Psychiatry 46:394–5, 2000
3. Lam RW, Kennedy SH: Using meta-analysis to evaluate evidence: practical tips and traps. Can J Psychiatry 50:167–74, 2005
4. Levine R, Fink M: The case against evidence-based principles in psychiatry. Med Hypotheses 67:401–10, 2006
5. Maier T: Evidence-based psychiatry: understanding the limitations of a method. J Eval Clin Pract 12:325–9, 2006
6. Slade M, Cahill S, Kelsey W, et al: Threshold 3: the feasibility of the threshold assessment grid (TAG) for routine assessment of the severity of mental health problems. Soc Psychiatry Psychiatr Epidemiol 36:516–21, 2001
7. Aarons GA: Mental health provider attitudes toward adoption of evidence-based practice: the Evidence-Based Practice Attitude Scale (EBPAS). Ment Health Serv Res 6:61–74, 2004
8. Aarons GA, McDonald EJ, Sheehan AK, et al: Confirmatory factor analysis of the evidence-based practice attitude scale in a geographically diverse sample of community mental health providers. Adm Policy Ment Health 34:465–9, 2007
9. Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993)
10. Frye v. United States, 293 F. 1013 (D.C. Cir. 1923)
11. Popper K: The Logic of Scientific Discovery. New York: Basic Books, 1959
12. General Electric Co. v. Joiner, 522 U.S. 136 (1997)
13. Kumho Tire Co., Ltd. v. Carmichael, 526 U.S. 137 (1999)
14. Rogers R, Shuman DW: Fundamentals of Forensic Practice: Mental Health and Criminal Law. New York: Springer, 2005
15. Welch C: Flexible standards, deferential review: Daubert's legacy of confusion. Harv J L Public Policy 29:1085–105, 2006
16. Federal Judicial Center: Reference Manual on Scientific Evidence (ed 2). Washington, DC: Federal Judicial Center, 2000
17. Gatowski SI, Dobbin SA, Richardson JT, et al: Asking the gatekeepers: a national survey of judges on judging expert evidence in a post-Daubert world. Law Hum Behav 25:433–58, 2001
18. Grove WM, Barden RC, Garb HN, et al: Failure of Rorschach-comprehensive-system-based testimony to be admissible under the Daubert-Joiner-Kumho standard. Psychol Public Policy Law 8:216–34, 2002

19. Ritzler B, Erard R, Pettigrew G: A final reply to Grove and Barden: the relevance of the Rorschach comprehensive system for expert testimony. Psychol Public Policy Law 8:235–46, 2002

20. Rogers R, Salekin RT, Sewell KW: Validation of the Millon Clinical Multiaxial Inventory for Axis II disorders: does it meet the Daubert standard? Law Hum Behav 23:425–43, 1999

21. Dyer FJ, McCann JT: The Millon clinical inventories, research critical of their forensic application and Daubert criteria. Law Hum Behav 24:487–97, 2000

22. Rogers R, Salekin RT, Sewell KW: The Millon Clinical Multiaxial Inventory: separating rhetoric from reality. Law Hum Behav 24:501–6, 2000

23. Rogers R, Jordan MJ, Harrison KS: A critical review of published competency-to-confess measures. Law Hum Behav 28:707–18, 2004

24. Grisso T: Reply to a critical review of published competency-to-confess measures. Law Hum Behav 28:719–24, 2004

25. Rogers R, Shuman DW: The Mental State at the Time of the Offense measure: its validation and admissibility under Daubert. J Am Acad Psychiatry Law 28:23–8, 2000

26. Poythress N, Melton GB, Petrila J, *et al*: Commentary on the Mental State at the Time of the Offense measure. J Am Acad Psychiatry Law 28:29–32, 2000

27. Kelly RF, Ramsey SH: Assessing and communicating social science information in family and child judicial settings: standards for judges and allied professionals. Fam Court Rev 45:22–41, 2007

28. Rogers R: Rogers Criminal Responsibility Assessment Scales (R-CRAS) and Test Manual. Odessa, FL: Psychological Assessment Resources, 1984

29. Rogers R, Bagby RM, Dickens SE: Structured Interview of Reported Symptoms (SIRS) and Professional Manual. Odessa, FL: Psychological Assessment Resources, 1992

30. Rogers R, Tillbrook CE, Sewell KW: Evaluation of Competency to Stand Trial-Revised (ECST-R) and Professional Manual. Odessa, FL: Psychological Assessment Resources, 2004

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

32. Melton GB, Petrila J, Poythress NG, *et al*: Psychological Evaluations for the Courts (ed 3). New York: Guilford Press, 2007

33. American Psychiatric Association: Psychiatric Services in Jails and Prisons (ed 2). Washington, DC: American Psychiatric Association, 2002

34. Robey A: Criteria for competency to stand trial: a checklist for psychiatrists. Am J Psychiatry 122:616–23, 1965

35. Lipsitt PD, Lelos D, McGarry AL: Competency for trial: a screening instrument. J Psychiatry 128:105–9, 1971

36. Laboratory of Community Psychiatry, Harvard Medical School: Competency to Stand Trial and Mental Illness (DHEW Pub. No. ADM-77-103). Rockville, MD: Department of Health, Education, and Welfare, 1973

37. Mossman D, Noffsinger SG, Ash P, *et al*: Practice guideline of the forensic psychiatric evaluation of competence to stand trial. J Am Acad Psychiatry Law 35(Suppl):S3–72, 2007

38. Heilbrun K, Rogers R, Otto RK. Forensic assessment: current status and future directions, in Taking Psychology and Law Into the Twenty-First Century. Edited by Ogloff JRP. New York: Kluwer, 2002, pp 120–46

39. Grisso T: Evaluating Competencies: Forensic Assessments and Instruments (ed 2). New York: Kluwer Academic, 2003

40. Poythress NG, Nicholson R, Otto RK, *et al*: Professional Manual for the MacArthur Competence Assessment Tool-Criminal Adjudication (MacCAT-CA). Odessa, FL: Psychological Assessment Resources, 1999

41. Everington C, Luckasson R: Manual for Competence Assessment for Standing Trial for Defendants with Mental Retardation: CAST-MR. Worthington, OH: IDS Publishing, 1992

42. Hoge SK, Bonnie RJ, Poythress N, *et al*: The MacArthur adjudicative competence study: development and validation of a research instrument. Law Human Behav 21:141–82, 1997

43. Miller GA: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 63:81–97, 1956

44. Baddeley A: The magical number seven: still magic after all these years? Psychol Rev 101:353–6, 1994

45. Cohen DA: Review of the Evaluation of Competency to Stand Trial-Revised. Mental Measurements Yearbook, Lincoln, NE, Accession No. 16093149, 2004. Available at http://libproxy.library.unt.edu:2055/ehost/detail?vid=3&hid=102&sid=bdf72205-c12f-4c62-aa8f-e6eedlalb8e2%40sessionmgr110&bdata=JnNpdGU9ZWhvc3QtbG12ZSZzY29wZT1zaXR1#db=loh&AN=16093149. Accessed January 20, 2008

46. Leark RA: Review of the Evaluation of Competency to Stand Trial-Revised. Mental Measurements Yearbook, Lincoln, NE, Accession No. 16093149, 2004. Available at http://libproxy.library.unt.edu:2055/ehost/detail?vid=3&hid=102&sid=bdf72205-c12f-4c62-aa8f-e6eedlalb8e2%40sessionmgr110&bdata=JnNpdGU9ZWhvc3QtbG12ZSZzY29wZT1zaXR1#db=loh&AN=16093149. Accessed January 20, 2008

47. Rush AJ, Pincus HA, First MB, *et al*: Handbook of Psychiatric Measures. Washington, DC: American Psychiatric Press, 2000

48. Gutheil TG, Bursztajn H: Avoiding ipse dixit mislabeling: post-Daubert approaches to expert clinical opinions. J Am Acad Psychiatry Law 31:205–10, 2003