The capacity to consent to treatment and research: A review of standardized assessment tools

Edward D. Sturman

Graduate Psychology, York University, BSB, 2nd floor, 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3

Received 12 May 2004; received in revised form 14 February 2005; accepted 20 April 2005

Abstract

The use of standardized tools for the assessment of competency is vital, given that informal assessments performed by physicians are idiosyncratic and unreliable [Marson, D. C., McInturff, B., Hawkins, L., Bartolucci, A., Harrell, L. E. (1997). Consistency of physician judgments of capacity to consent in mild Alzheimer’s disease. Journal of the American Geriatrics Society, 45, 453–457]. A number of instruments have been developed for this purpose, which are outlined in the following review. For the most part, these tools rest on a definition of competency put forward by Appelbaum and Roth (1982) [Appelbaum, P. S., & Roth, L. H. (1982). Competency to consent to research. A psychiatric overview. Archives of General Psychiatry, 39, 951–958], which incorporates various legal standards. Therefore, the present review will summarize current thinking on competence as well as the strengths, weaknesses, and psychometric properties of existing measures. As competency assessment will be more or less necessary depending on the patient population, another goal of the paper was to assemble the major findings pertaining to patient groups that have impairments in capacity. Although competency varies between and within diagnostic groups, empirical studies consistently demonstrate increased risk for some populations, and clinicians or researchers should be aware of this information. As such, this review may be useful as an initial source for interested researchers or clinicians.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Competency; Informed consent; Measurement

E-mail address: sturman@yorku.ca.

0272-7358/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
1. Introduction

Contemporary ideas on informed consent stem, in part, from the Nuremberg Code following the Second World War. The code specified that consent to research should be voluntary and that participants should have free choice, adequate understanding, and the capacity to give consent. Currently, if a person fails to have the capacity to consent to treatment or research, substitute decision makers are employed (e.g., legal representatives or family members). Thus, a judgement by health care professionals that a person is not competent to provide informed consent may have a profound effect on the autonomy, or free will, of an individual. But what constitutes the capacity to provide informed consent and how is it assessed?

2. Defining competency

Before a definition is offered, it should be noted that the terms “capacity,” “legal standards,” “abilities,” and “competency” are often used interchangeably to describe the same phenomena and yet these terms have different meanings. Marson (2001) has differentiated capacity and competency as such: “capacity denotes a clinical status as judged by a health care professional whereas competency denotes a legal status as judged by a legal professional, i.e., a judge.” Yet, despite the fact that competency is a legal status, physicians and psychiatrists are frequently required to make competency judgements in the course of their duties (Marson, Bronwyn, Hawkins, Bartolucci, & Harrell, 1997). This review will use the term competence to refer to a judgement, either by a health care professional or legal professional, as to whether a person is able to provide informed consent. Capacity will refer to the resources that a patient or participant can rely upon to provide informed consent and is thought to lie on a continuum. Finally, abilities are treated as task-specific dimensions, which inform legal standards of competency.

The most common legal standards for competency were initially conceptualized as “tests” by Roth, Meisel, and Lidz (1977). According to Roth et al. (1977), competency could be demonstrated by the following tests: (1) evidencing a choice; (2) reasonable outcome of choice; (3) choice based on rational reasons; (4) ability to understand; and (5) actual understanding. Based on this original work, Appelbaum and Roth (1982) outlined four legal standards, which have come to be used by the majority of researchers in the field. These standards, which correspond to functional abilities, were as follows: (1) evidencing a choice; (2) factual understanding of the issues; (3) rational manipulation of information; and (4) appreciation of the nature of the situation.

Evidencing a choice is the least stringent standard and involves the simple communication of consent or lack thereof. It is often implied and not stated explicitly in American courts. Although the standard is easy to meet, the presence of certain psychiatric conditions may make it more difficult (e.g., mutism due to catatonia or severe depression, mania, thought disorder stemming from psychosis, etc.; Appelbaum & Roth, 1982). According to Appelbaum and Roth (1982), factual understanding of the proposed research/treatment may be the most widely accepted standard of competency, combining the ability to understand and actual understanding from Roth et al. (the ability to understand refers to the patient’s capacity to understand the risks and benefits of treatment as well as possible alternatives). The third standard, the rational manipulation of information, focuses on the manner in which participants utilize information to arrive at a decision. Psychiatric factors that may adversely affect performance include delusions, hallucinations, thought disorder, panic, anxiety, depression, and euphoria. However, this ability cannot
be ascertained on the basis of psychiatric factors alone because they may or may not affect global decision making. Appelbaum and Roth (1982) claimed that an appreciation of the nature of a situation was the strictest standard. Generally, it is the ability of participants to relate factual understanding to their own situation. Appelbaum and Grisso (1995) later separated this concept into a lack of insight or awareness of one’s psychiatric illness and the inability to appreciate the risks and benefits of treatment.

Appelbaum and Roth (1982) suggested that the standards that will actually be used are the product of competing societal values and the goals of the researchers. For instance, encouraging autonomy for the individual will entail the use of a relatively lax standard for competency. On the other hand, beneficence, defined as “the ethical imperative to protect others from harm and to do good whenever possible” (p. 956), requires a strict set of standards, such as rationality or appreciation. Appelbaum and Roth also pointed to temporal considerations whereby competency should be reassessed as clinical status changes over time. For example, when a patient enters a study, a fairly low standard for competency may be used but, as his/her clinical picture improves, competency should be reassessed using a tougher standard.

3. The measurement of abilities related to competency

This review will outline standardized instruments that measure one or more functional abilities related to competence (evidencing a choice, understanding, reasoning, and appreciation). It should be noted that some of the test developers do not refer to specific abilities but instead view their instrument as measuring competency. However, the determination of competency requires a proper assessment that incorporates many factors relevant to the unique circumstances of the patient and cannot be ascertained on the basis of test scores alone. For instance, Grisso and Appelbaum (1998) have suggested that, in addition to functional abilities, a determination of competence or incompetence must take into account the decision-making demands placed on a patient, which include different situational and social factors, as well as the consequences of a judgement in terms of beneficence and autonomy. For the present review, attention will be given to the format, reliability, validity, and the type of population that each tool has been tested on (also see Table 1).

3.1. Hopkins Competency Assessment Test (HCAT)

The HCAT was developed to aid clinicians in “forming an opinion” regarding the competency of a patient (Janofsky, McCarthy, & Folstein, 1992). The HCAT consists of an essay explaining the informed consent process and power of attorney (written at different grade levels). There are six questions regarding this essay in a true/false and sentence completion format. Scores range from 0 to 10, with higher scores indicating greater competence. An advantage of the HCAT is that it is quick, requiring only 10 min to administer, and it is straightforward. Furthermore, it can be administered by non-clinicians, increasing its potential use as a screening instrument for large numbers of patients or subjects. However, it only assesses the understanding of patients and, as such, is limited.

Janofsky et al. (1992) tested the HCAT in a mixed sample of 16 medical inpatients and 25 adult/geriatric patients from a psychiatric ward. Interobserver reliability for the measure was .95 between two raters. Scores on the HCAT were compared to the judgement of a forensic psychiatrist with experience in competency assessment. Using a cutoff score of 3 or less to indicate incompetence, the HCAT demonstrated 100% sensitivity and 100% specificity.
Barton, Mallik, Orr, and Janofsky (1996) used the HCAT in a sample of 44 nursing home admissions and found high interobserver reliability between three raters, ranging from .96 to .99. With the HCAT as a gold standard, clinicians identified as incompetent only 65% of those subjects who scored in the incompetent range. One possibility is that clinicians were attempting to grant as much autonomy as possible to patients. Conversely, clinicians identified as competent 92% of those patients that fell in the competent range on the HCAT. The high sensitivity may have been due to the fact that patients were presumed to be competent. Overall, the HCAT identified 45% of nursing home admissions as incompetent and 55% as competent.

Silberfeld, Stephens, and O’Rourke (1994) found that for Mini Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975) scores outside the range of 18–24, the HCAT had low sensitivity or a higher chance of Type I error. Of course, the MMSE was not designed as a capacity instrument and is therefore problematic as a standard. More recently, one group of researchers demonstrated the utility of the HCAT in assessing changes in understanding among patients with schizophrenia (Palmer, Nayak, Dunn, Appelbaum, & Jeste, 2002).

3.2. Competency Interview Schedule (CIS)

The CIS (Bean, Nishisato, Rector, & Glancy, 1994) has the advantage of covering all four abilities pertaining to competency (Appelbaum & Roth, 1982) in a 15-item structured interview format. Each item is rated on a seven-point Likert scale, with lower numbers indicating greater capacity. The developers of the CIS needed a test of competency for patients facing electroconvulsive therapy (ECT) in Ontario. The CIS was tested in 96 psychiatric inpatients (a mixed sample of patients with schizophrenia, schizoaffective disorder, depression, and mania). The internal consistency was excellent (Cronbach’s $\alpha$ was .96) and interrater reliability was also very good. Test–retest reliability over 24 h was good for some items but not others. However, there were problems with this statistic for several reasons: patients needed to be retested quickly, meaning that practise effects could occur; only 13 participants were retested; and only those cases in which competency was uncertain were retested. The CIS agreed with the judgement of an attending physician, thus lending it criterion validity. All of those patients deemed competent by the physician had significantly lower scores for every item compared to those deemed incompetent. A factor analysis revealed only one factor in the CIS despite the subsets of items corresponding to the four abilities. The developers referred to the factor as “general competency.”

3.3. MacArthur treatment competence study

The MacArthur study (1995; I Appelbaum and Grisso; II Grisso, Appelbaum, Mulvey, and Fletcher; III Grisso and Appelbaum) was geared towards developing standardized tools to assess all four functional abilities related to competence (Appelbaum & Roth, 1982). Separate instruments were created to measure each of the abilities and were tested in three groups of inpatients: schizophrenic/schizoaffective ($n=75$), depressed ($n=92$), and ischemic heart disease ($n=82$). Although the content of the instruments was customized for each diagnosis, all of the measures were standardized to the extent that comparisons within and across groups were possible. Objective and reliable scoring criteria were also devised for each measure (Appelbaum & Grisso, 1995).

The Understanding Treatment Disclosure (UTD) provided a five-paragraph disclosure like that used in consent forms and came in three different versions for each group. Understanding was assessed either
<table>
<thead>
<tr>
<th>Instrument</th>
<th>Component(s) measured</th>
<th>Type of measure</th>
<th>Length</th>
<th>Psychometric properties</th>
<th>Types of samples</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCAT*</td>
<td>Understanding/ comprehension</td>
<td>Self-report</td>
<td>10 min</td>
<td>High interrater reliability</td>
<td>Mixed (medical and general/geriatric psychiatric) ((n = 16 \text{ and } n = 25))</td>
<td>Good agreement with physician judgements and can detect changes in understanding over time</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>High sensitivity/specificity in some samples</td>
<td>Nursing home ((n = 44))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Alzheimer’s disease ((n = 16))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Schizophrenia ((n = 30; n = 25))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HIV ((n = 25))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Alzheimer’s disease ((n = 37))</td>
<td></td>
</tr>
<tr>
<td>MacCAT-CR*</td>
<td>Choice, understanding, appreciation, reasoning</td>
<td>Structured interview</td>
<td>15–20 min</td>
<td>High interrater reliability</td>
<td>Depression ((n = 26))</td>
<td>Can be adapted for any research design</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Possibly low test–retest reliability (ceiling effect)</td>
<td>Schizophrenia ((n = 30; n = 25))</td>
<td>Correlations with BPRS and MMSE; distinguishes between patients and control subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cutoffs for capacity have been used for each component</td>
<td>HIV ((n = 25))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Alzheimer’s disease ((n = 37))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dementia ((n = 31))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Depression ((n = 35))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anorexia nervosa ((n = 10))</td>
<td></td>
</tr>
<tr>
<td>MacCAT-T*</td>
<td>Choice, understanding, appreciation, reasoning</td>
<td>Semi-structured interview</td>
<td>15–20 min</td>
<td>High interrater reliability</td>
<td>Schizophrenia ((n = 40; n = 43)) and older adults with psychosis ((n = 16))</td>
<td>Utilizes patient chart</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dementia ((n = 31))</td>
<td>Correlations with individual BPRS items and HCAT; moderate agreement with physician ratings and distinguishes diagnostic groups</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Depression ((n = 35))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anorexia nervosa ((n = 10))</td>
<td></td>
</tr>
<tr>
<td>CCTI*</td>
<td>Choice, understanding, appreciation, reasoning, reasonable decision</td>
<td>Semi-structured interview</td>
<td>Vignettes require 20–25 min to present</td>
<td>High interrater reliability for all standards</td>
<td>Alzheimer’s disease ((n = 29 \text{ mild/moderate}; n = 29 \text{ mild}; n = 21 \text{ mild/moderate}))</td>
<td>Well validated instrument</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Parkinson’s disease ((n = 20))</td>
<td>Presents hypothetical vignettes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Number of items unknown</td>
<td>Cutoffs have been used for each standard</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Domain</th>
<th>Format</th>
<th>Duration</th>
<th>Reliability/Validity</th>
<th>Population</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESC&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Understanding</td>
<td>Questionnaire</td>
<td>Very short (five items)</td>
<td>Unknown (not published)</td>
<td>Schizophrenia ($n=25$)</td>
<td>Can be customized for each study</td>
</tr>
<tr>
<td>ACE&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Understanding, appreciation</td>
<td>Semi-structured interview</td>
<td>15 min</td>
<td>High interrater reliability</td>
<td>Medical inpatients ($n=100$)</td>
<td>Tailored to each individual’s diagnosis and treatment</td>
</tr>
<tr>
<td>CSA&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Appreciation</td>
<td>Semi-structured interview</td>
<td>10–20 min</td>
<td>High internal reliability ($r$ ranges from .83 to .89)</td>
<td>Schizophrenia ($n=39$)</td>
<td>Able to distinguish normal from schizophrenic groups</td>
</tr>
<tr>
<td>CAT&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Choice, understanding, appreciation, reasoning</td>
<td>Semi-structured interview</td>
<td>Unknown duration</td>
<td>No info on reliability</td>
<td>Primary care patients ($n=20$)</td>
<td>Good agreement with independent psychiatric assessment</td>
</tr>
<tr>
<td>CIS&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Choice, understanding, appreciation, reasoning</td>
<td>Structured interview</td>
<td>Unknown duration</td>
<td>High interrater reliability and internal consistency</td>
<td>Psychiatric inpatients ($n=96$)</td>
<td>Able to distinguish competent and incompetent patients as determined by attending physician</td>
</tr>
<tr>
<td>SICIATRI&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Choice, understanding, appreciation, reasoning</td>
<td>Structured interview</td>
<td>20 min</td>
<td>Inter-rater agreement good for most items, poor for others</td>
<td>Psychiatric ($n=25$) and medical ($n=23$) inpatients</td>
<td>Potentially useful in clinical settings</td>
</tr>
</tbody>
</table>

<sup>a</sup> HCAT=Hopkins Competence Assessment Tool; Mac-CAT-CR=MacArthur Competence Assessment Tool—Clinical Research; Mac-CAT-T=MacArthur Competence Assessment Tool—Treatment; CCTI=Capacity to Consent to Treatment Instrument; ESC=Evaluation to Sign Consent; ACE=Aid to Capacity Evaluation; CSA=California Scale of Appreciation; CAT=Capacity Assessment Tool; CIS=Competency Interview Schedule; SICIATRI=Structured Interview for Competency/Incompetency Assessment Testing and Ranking Inventory.
after each paragraph or the whole disclosure. Perceptions of disorder (POD) measured the appreciation of information related to subjects’ disorders and treatments. Accordingly, it was divided into the non-acknowledgement of disorder (NOD) and non-acknowledgement of treatment (NOT) subtests. It also came in three different versions for each diagnostic group. Thinking Rationally About Treatment (TRAT) evaluated the cognitive functions that are used to decide between alternative treatments. Examples of these functions include seeking information, consequential thinking, complex thinking, and probabilistic thinking. Expressing a choice of treatments was also assessed with the TRAT (Grisso & Appelbaum, 1995).

Interscorer reliability was very good for understanding (UTD) and reasoning (TRAT) between 10 raters and a “master scorer.” The subtests of the UTD1 and POD showed acceptable internal reliability using Cronbach’s $\alpha$. It should be noted that a different group of investigators had also found the UTD to be internally consistent and predictive of clinical assessments of competency (Pruchino, Smyer, Rose, & Hartman-Stein, 1995). Some TRAT subtests did not appear to be internally reliable and a separate, more internally consistent version of the test (TRAT-2) was created by excluding two cognitive functions. Test–retest correlations for the measures were generally between .50 and .80. For patients with schizophrenia, changes in symptom severity between the first and second tests were related to performance on the reasoning and understanding scales but not appreciation. For depressed patients, only appreciation was related to changes in symptom severity. As for correlations between tests, various aspects of understanding (UTD) and reasoning (TRAT-2) appeared to be related but neither measure was related to appreciation (POD). Altogether, the measures appeared to be reliable and valid for the purposes of the MacArthur study. However, the authors did not recommend the instruments for routine clinical use due to the time and effort necessary for administration and scoring (Grisso & Appelbaum, 1995).

3.4. MacArthur Competence Assessment Tool—Treatment (MacCAT-T)

Grisso, Appelbaum, and Hill-Fotouhi (1997) developed the MacCAT-T for use in clinical settings. The MacCAT-T is a semi-structured interview based on the aforementioned MacArthur Treatment Competence Research Instruments. These instruments, while reliable and valid, were time-consuming (requiring 60–90 min to administer) and it was not possible to use them in relation to the patient’s own symptoms and treatments. They also had overly complex scoring criteria.

The MacCAT-T utilizes information from each patient’s chart in order to make the test personally relevant. Patients receive either adequate, partial, or inadequate ratings for each item. There is a score for each ability (understanding, reasoning, and appreciation) but no overall score or cutoff scores for any of the abilities. The MacCAT-T is not designed to determine global competence but rather is intended to identify areas of relative capacity or incapacity and is to be interpreted in the context of other relevant clinical information. It is not time-consuming, requiring approximately 15–20 min to administer (Grisso et al., 1997). A manual is available for the MacCAT-T (Grisso et al., 1997) as well as a training video.

---

1 The UTD and its subtests had been discussed in preliminary reports under the title Measuring Understanding of Disclosure (MUD), but it appears to be the same test (Grisso & Appelbaum, 1991). A version developed for clinical use (MUD-VH) was found to be overly stringent by one group of researchers (Polythress, Cascardi, & Ritterband, 1996).
Grisso et al. (1997) tested the MacCAT-T in a sample of hospitalized patients with schizophrenia. High interrater reliability was demonstrated for each ability (.99 for understanding, .87 for appreciation, and .91 for reasoning). The schizophrenia group \( (n=40) \) received significantly worse scores for each ability compared to a matched community group \( (n=40) \). Total scores on the Brief Psychiatric Rating Scale (BPRS; Overall & Gorham, 1962) were not related to any ability but individual BPRS proved to be. For instance, understanding was negatively correlated with hallucinations, disorientation, and conceptual disorganization, among other BPRS items. Therefore, the MacCAT-T showed excellent reliability and promise as a valid measure of competency. The authors reported having no problem individualizing the MacCAT-T for each patient. The instrument was thought to be most practical for cases of ambiguous competency (some cases are clearly not competent and do not require formal assessment). Therefore, the authors believe it will be of greatest use in cases involving legal proceedings to determine competency and for psychiatrists conducting formal assessments.

Vollman, Bauer, Danker-Hopfe, and Helmchen (2003) tested the MacCAT-T in a German sample composed of patients with dementia \( (n=31) \), schizophrenia \( (n=43) \), and depression \( (n=35) \). The dementia and schizophrenia groups showed the greatest degree of impairment on each ability as might be expected from previous findings. Judgements of competency by an attending physician generally showed moderate associations with the MacCAT-T for the schizophrenia and dementia groups (the depressed group did not have sufficient impairment to allow for proper comparisons between the two methods).

Recently, the MacCAT-T and HCAT have also been used in an elderly population (Palmer, Nayak, Dunn, Appelbaum, & Jeste, 2002). This research group assessed 16 middle-aged and older adults with psychosis in what was acknowledged to be a preliminary study. Nevertheless, a wide range of MacCAT-T scores was obtained and the pattern of scores was similar to the original validation study. Intraclass correlation coefficients between different raters were: .85 for understanding, .87 for appreciation, and .75 for reasoning. The HCAT was used in this study to assess changes in understanding over three trials. In fact, scores did significantly improve on the HCAT and the authors acknowledged its use in measuring changes in understanding. However, they also noted that the HCAT is “much less comprehensive than the MacCAT-T and does not actually require the patient to take a decision-making perspective.”

The MacCAT-T may not be a valid instrument in every population. Tan, Hope, and Stewart (2003) found that 10 patients with anorexia nervosa rated very highly on all abilities (comparable to control groups that had been used in previous studies; e.g., Grisso et al., 1997). However, a qualitative analysis indicated that these patients did indeed have problems with decision making in regards to treatment. Themes emerged relating to alterations in attitudes towards death; anorexia as a part of personal identity; and ambivalence in regards to treatment or even successful treatment. The investigators noted that the MacCAT-T procedure of going over the facts of the disorder and treatment was “patronizing” towards this group of participants (who were already well-informed) and “awkward” for the researchers.

3.5. MacArthur Competence Assessment Tool—Clinical Research (MacCAT-CR)

Grisso and Appelbaum (1998; as cited in Appelbaum, Grisso, Frank, O’Donnell, & Kupfer, 1999) developed a separate version of the MacCAT for assessing competency in clinical research, the MacCAT-CR. Like the MacCAT-T, it is a structured interview, requiring approximately 15–20 min, and
measures the same abilities. It can also be customized to each research protocol and diagnostic sample. A training manual for the MacCAT-CR is widely available, which provides information on the manner in which it can be customized to assess each ability (Appelbaum & Grisso, 1996).

Appelbaum et al. (1999) used the MacCAT-CR in a sample of 26 depressed female outpatients but the psychometric properties of the instrument could not be properly evaluated. Test–retest reliability was low, probably due to ceiling effects (depressed subjects scored high on each component at both test administrations). Subsequent studies have found interrater agreements to be excellent (Carpenter et al., 2000; Karlawish, Casaret, & James, 2002; Kim, Caine, Currier, Leibovici, & Ryan, 2001). To date, the MacCAT-CR has been used to assess competency in a variety of disorders such as depression (Appelbaum et al., 1999), dementia (Karlawish et al., 2002; Kim et al., 2001), schizophrenia (Carpenter et al., 2000), and cancer (Casaret, Karlawish, & Hirschman, 2003).

Carpenter et al. (2000) evaluated the MacCAT-CR in a sample of 30 inpatients and outpatients with schizophrenia. They obtained some validity for the measure as BPRS total scores were negatively correlated with the understanding and reasoning scales. Two separate groups of researchers adapted the MacCAT-CR for use in patients with Alzheimer’s disease who were asked to participate in hypothetical clinical trials (Karlawish et al., 2002; Kim et al., 2001). Both groups of researchers were able to establish cutoffs for determining competency on each ability. Kim et al. (2001) accomplished this by using expert opinion and normative data, while Karlawish et al. (2002) defined scores that were two standard deviations below a control group as poor. Karlawish et al. demonstrated the validity of the measure by clearly distinguishing patients from normal control subjects and by relating the understanding component to MMSE scores. However, in both studies, experts rated a substantially greater number of patients as competent compared to the MacCAT-CR. This raises the possibility that the MacCAT-CR may be overly stringent in this population. In general, the MacCAT-CR appears to be a reliable and valid instrument whose greatest strength may be the flexibility to adapt to almost any research design with any diagnostic group.

### 3.6. Structured Interview for Competency/Incompetency Assessment Testing and Ranking Inventory (SICIATRI)

The SICIATRI was designed for use in a variety of clinical settings such as admission, medication, operation, or examination (Tomoda et al., 1997). It was a valuable measure at the time of its development because it covered all of the major abilities related to competency proposed by Appelbaum and Roth (1982) and could be tailored to each individual patient. However, with the publication of the MacCAT-T, it no longer represents the only option for clinical settings.

The SICIATRI is a 12-item structured interview and each question is rated on a three-point scale. For each item, the authors have developed explanations, probes, and anchor points with definitions. Tomoda et al. (1997) examined the psychometric properties of the SICIATRI in a sample of 25 psychiatric and 23 medical inpatients. They found good interrater agreement for the majority of items, although some items had low \( \kappa \) values.

Once patients have been scored on the 12 items, they can be ranked in terms of their total competency with the Ranking Inventory for Competency. This measure consists of five levels (from 0 to 4), with higher levels indicating greater competency on all of the components. Compared to a “Global Assessment of Competency” by an attending physician, the Ranking Inventory for Competency displayed .83 sensitivity and .67 specificity. A subsequent report of this same sample revealed that
psychiatric patients differed significantly from medical patients in their frequency of competent responses to only two SICIATRI items (Kitamura et al., 1998).

3.7. Capacity to Consent to Treatment Instrument (CCTI)

Marson, Ingram, Cody, and Harrell (1995) developed the CCTI for use in patients with Alzheimer’s disease (AD). This instrument utilizes two clinical vignettes described only as A, neoplasm, and B, cardiac. Each vignette depicts a particular medical problem with associated symptoms as well as the risks/benefits for two treatment alternatives. Competency is assessed in relation to five legal standards. These standards are essentially the same as those suggested by Appelbaum and Roth (1982), except for one additional standard pertaining to a reasonable choice of treatments (recall that this had been included in the formulation by Roth et al., 1977). The legal standards (LS) in order of least to most stringent were: LS 1, evidencing a choice; LS 2, making a reasonable decision; LS 3, appreciation of the consequences of a treatment choice; LS 4, providing rational reasons for a treatment choice; and LS 5, understanding of the treatment situation and choices.

Some construct validity was obtained for the measure by comparing the mild ($n=15$) and moderate ($n=14$) AD groups to a normal, elderly control group (Marson et al., 1995). For the more stringent standards, the control group performed significantly better than both AD groups and appeared to have very good capacity for consent. In contrast, the mild and moderate AD patients showed high levels of incapacity for the tougher standards.

Marson et al. (2000) speculated that by using a common set of legal standards, physicians would display more accurate and consistent judgements of competency for people with Alzheimer’s disease. A previous study (Marson et al., 1997) had demonstrated a lack of agreement between physicians for judgements of competency in a group of patients with mild AD (the agreement was only 56%). After training five physicians in the use of the CCTI, a much higher level of concordance was indeed obtained. Judgements of overall competency for an AD group ($n=21$) rose to 76% agreement between the five physicians. One implication of the findings is that physicians will be better able to make reliable judgements of competency by using standardized measures such as the CCTI. Second, the value of formally training physicians in competency assessment is evident. One limitation of the CCTI is the use of vignettes, which may not elicit the same responses as disclosures related to a personal medical problem. Marson et al. feel that this potential limitation is offset by the value of a standardized instrument.

3.8. Evaluation to Sign Consent (ESC)

The ESC was developed by Love in 1988 (unpublished; as cited in Deronzo, Conley, & Love, 1998). The ESC assesses the factual understanding of subjects and is composed of only five items. Its greatest strength is that it can be adapted to any research design with ease. The ESC is therefore short, concrete, and only asks about facts relevant to the study. Deronzo et al. (1998) believe that the ESC may not be sufficiently rigorous in contrast to the MacCAT-CR, which may be overly difficult.

Moser et al. (2002) used both the ESC and MacCAT-CR in a study assessing the competence of patients with schizophrenia ($n=25$) and HIV ($n=25$). The ESC was able to identify those individuals who displayed poor understanding on the MacCAT-CR, raising the question of whether the ESC might be a preferable instrument as it is much briefer and does not require as much effort to customize for each
study. However, Moser et al. point out that the ESC does not evaluate appreciation, reasoning, or the capacity to evidence a choice. Therefore, the MacCAT-CR is better suited to reveal specific deficiencies in a patient’s capacity, which may then be addressed by researchers or health care practitioners.

3.9. The Mini Mental Status Examination (MMSE)

Etchells, Schuchman, Workman, and Craven (1997) claim that clinicians often use informal clinical impressions or MMSE scores to arrive at decisions of competency for major medical treatment. Using a psychiatric assessment of competency as a gold standard, these researchers found that clinical impressions and SMMSE scores were rather inaccurate. Reducing or raising the cutoff on the SMMSE to 16 and 26 would correctly classify most people but would falsely classify others. Kim and Caine (2002) looked at the usefulness of the MMSE to predict competency for research studies. With the MacCAT-CR as the gold standard, the MMSE was found to significantly add to the identification of incompetence in Alzheimer’s disease. Unfortunately, no MMSE cutoff score produced both high sensitivity and high specificity. They concluded that the MMSE was not a good predictor of incompetence, but by using two cutoff scores, it could still be useful. A cutoff score of 26 had a sensitivity of 91–100% in identifying incompetent patients while the cutoff score of 19 had a specificity of 85–94%.

Kim, Karlawish, and Caine (2002) examined other studies that had found similar results with scores of 16–20 and below correctly predicting incompetence, and scores 24–26 and above correctly identifying competence. Scores that fall in between these values are problematic, but Kim et al. argue that the test is so widely applied in some populations, such as dementia, that it can be used as a relatively inexpensive source of information regarding competency. Thus, the MMSE should be viewed as a blunt instrument for ascertaining competency.

3.10. Aid to Capacity Evaluation (ACE)

Some newer scales for assessing capacity have shown promise but require further validation. The ACE assesses appreciation of disorder/treatment and understanding of informed consent (Etchells et al., 1999). Like the CIS, the test was developed in Ontario where these two standards are relevant to existing law. It is a semi-structured test that enables clinicians to rate patients as: definitely incapable, probably incapable, probably capable, and definitely capable. Some advantages of the ACE include a short administration time (approximately 15 min) and the fact that it is tailored to each patient’s disorder and treatment. In a sample of 100 medical inpatients with questionable capacity, Etchells et al. (1999) found high overall agreement between clinician/nurse ratings on the ACE and expert competency assessments. Agreement between the treating clinician and a research nurse was 93% for the ACE assessments (κ of .79).

3.11. California Scale of Appreciation (CSA)

The CSA was designed to measure only the appreciation component of Appelbaum and Roth’s (1982) formulation of competency (Saks et al., 2002). Lack of appreciation was measured in relation to a “patently false belief.” These are beliefs that violate the laws of nature, extremely improbable beliefs, or beliefs that distort facts. Saks et al. argue that using this standard for appreciation increases patient beneficence by protecting vulnerable individuals (those with patently false beliefs) while granting autonomy to patients as they may still have a wide range of beliefs. It also prevents discrimination of the
mentally ill as a whole because people are only deemed non-competent when they possess “major distortions.”

The CSA contains 13 close-ended questions and 6 open-ended questions that refer to a hypothetical consent form. It was found to have good reliability based on a sample of 39 middle-aged and older patients with Schizophrenia and 15 normal control subjects. Internal consistency, as measured by Cronbach’s α, ranged from .83 to .89, while interrater agreement, as measured by percent agreement and κ for each item, was also generally high. Test–retest reliability was 100%, but this was based on data from only six participants. The CSA identified all of the “normal” participants as competent for the appreciation component and 64.1–76.9% of the patient group. It was positively correlated with the dementia rating scale total score as well as the attention and conceptualization subscores. Some limitations were the small sample and consequent lack of power, the older nature of the sample, and the use of relatively stable patients with schizophrenia. Further work with the scale is required, but it is promising as a specific measure of appreciation.

3.12. Capacity Assessment Tool (CAT)

Other researchers have developed instruments for specific populations or circumstances. The CAT is a relatively short capacity measure designed for use in primary care settings (Carney, Neugroschl, Morrison, Marin, & Siu, 2001). The CAT is administered in regards to a specific real-world decision-making task that the patient faces. It covers four major abilities pertaining to competency (Appelbaum & Roth, 1982) and it is possible to obtain an overall score indicating each patient’s capacity in relation to the specific decision at hand.

The CAT was tested on 20 participants (median age of 82 years) faced with major medical treatments or invasive procedures that required informed consent. Assessments conducted with the CAT showed very high agreement with an independent psychiatric assessment for each ability and for overall competency. However, there are no data concerning the reliability of this measure and further testing is certainly required.

Baergen (2002) has noted several advantages of the CAT such as its focus on capacity for specific health care decisions (not global determinations of competency) and its potential use in many different groups of patients. In comparing it to the MacCAT, he judged it to have more structure, clearer scoring procedures, and thresholds that can be used to determine competence.

3.13. Limitations of existing standardized measures

Baergen (2002) has also outlined several limitations of the CAT, which pertain to other competency measures as well. First, in terms of the understanding component, patients may repeat or paraphrase details about their treatment even though they have little real understanding. If they lack scientific or medical knowledge, then misunderstandings or mistaken beliefs are possible. Second, the reasoning component requires patients to recall the mental processes they used in arriving at a decision but these types of self-reports have been found to be inaccurate and patients will not really be able to perform this task (although it may provide other valuable information such as knowledge of important details that should go into a decision). Third, patients who perform poorly in terms of the abilities measured do not necessarily lack these same abilities. Performance could be affected by numerous factors such as lack of motivation, inattention, mistrust, or a misunderstanding of expectations. Finally, a savvy patient can tell
investigators what he/she wants to hear, despite having a markedly different reasoning process (e.g., anorexic patients).

While these points may be valid, and raise issues that must be addressed, they do not necessarily undermine the use of standardized measures. It is important to remember that these measures do not constitute a determination of competency, but are rather one part of the assessment. Specifically, the different instruments attempt only to quantify functional abilities that are relevant to competence. A proper assessment should reveal tendencies to paraphrase; in fact, some measurement tools such as the MacCAT-T test patients’ understanding by asking them to restate the relevant facts in their own words. Likewise, the assessment should also take into account an apparent lack of motivation, inattention, mistrust, or misunderstanding. The clinician may be able to remedy these problems to some extent before testing patients again. Otherwise, they may play a role in his/her judgements. Either way, it is clear that measurement should be distinguished from assessment.

Baergen’s assertion that savvy individuals will be able to manipulate investigators is a point that deserves serious consideration. What seems to be at issue are “patently false beliefs,” held by patients who are motivated to hide these from health care professionals and have the presence of mind to do so. Thus, it is the appreciation component that is problematic. Measures of understanding and reasoning do not suffer this limitation as they are not reliant on self-report but are rather tests of patients’ abilities. Unfortunately, aside from the CSA, existing instruments are ill equipped to deal with this issue. Recall that anorexic patients scored highly on all aspects of the MacCAT-T despite the fact that a qualitative analysis revealed ambivalence over treatment. It would therefore be possible for a grossly underweight anorexic patient to be fully capable of refusing treatment, despite having distorted or false beliefs about their self-image and the necessity of treatment. It should be noted that this patient would likely have adequate capacity on every component, save for appreciation, but this one deficit could be hidden and have a profound impact on the choice of the patient. This is a limitation that holds little promise of being solved because self-report measures, even diagnostic interviews, are easily manipulated by clever and motivated individuals. However, in the majority of cases, patients will either not possess patently false beliefs that relate to their appreciation of disorder and treatment, or their ego function will be too impaired to hide such beliefs. Thus, existing instruments for measuring competency remain useful despite this limitation, except possibly in the anorexic population.

The most serious limitation of the various competency instruments lies in their validity testing. Many of the instruments have been tested only once, despite showing promise as a useful measure. This may be due to the emergence of the MacCAT instruments as the gold standard, rendering some other instruments obsolete. In other cases, such as the CAT or CSA, the measures have only been developed recently and more data may be available in the future.

Most commonly, validity has been established by comparing the various instruments with expert psychiatric opinion, which is understandable given the absence of a gold standard until fairly recently. However, psychiatrists’ assessments may be idiosyncratic and pose a problem for the reliability of findings. The ESC and HCAT (which measure understanding) have been compared to the MacCAT-CR and MacCAT-T, respectively, and have performed well. Thus, concurrent validity has been demonstrated for some instruments, but only for understanding.

Another issue uncovered in this review is the small sample sizes that have been used to test the respective interviews. Most of the sample sizes are adequate for pilot studies, but are insufficient to properly evaluate an instrument. A quick look at Table 1 reveals that only the CIS and ACE used samples approaching 100 participants. Also, the majority of interviews have been tested in very specific
populations and may, indeed, be intended only for use with certain types of disorder. Only the MacCAT instruments have been tested in a variety of different populations, and appear to be flexible for use in any patient group.

It should also be noted that despite the availability of standardized instruments, many investigators continue to develop their own proxy measures. Kim et al. (2002) reviewed studies on competency in cognitively impaired elderly patients and found that “with few exceptions, every research group developed its own instrument for measuring decisional abilities.” Of 28 separate study samples, four assessed competency with general clinical impression or interview, while the other 24 used a total of 18 different tools for measuring competency. Many of the proxy measures incorporated vignettes that described hypothetical scenarios. Comparisons across studies were further complicated by the fact that some studies examined one or two abilities related to competency, while others focussed on completely different abilities or all four (evidencing a choice, understanding, reasoning, and appreciation).

Such disparity in definitions of competency is due, in part, to differing legal formulations from one jurisdiction to another. However, the four overarching components of Roth et al. (1977) and Appelbaum and Roth (1982) are generally agreed upon by most researchers today. Thus, many of the measures are overinclusive in the definition and measurement of competency compared to the legislation for any one region and provide researchers or clinicians with a safe bet.2

3.14. Reliability

Virtually all of the instruments covered in the present review have demonstrated good interrater reliability. The CSA and CIS have additionally proven to have high internal reliability. However, only the MacCAT-CR, CIS, and CSA have been evaluated for test–retest reliability. For the MacCAT-CR, a ceiling effect may have been operating, whereas the CIS demonstrated only adequate test–retest reliability for some items but not others. Sample sizes were too small in each case to properly evaluate the test–retest reliability. Likewise, the CSA demonstrated perfect test–retest reliability, but this finding was based on only six cases. It should be noted that test–retest reliability might not be an appropriate statistic, given that performance on some abilities varies with symptom severity (Grisso et al., 1995).

4. Competency in specific populations

The capacity to consent to research or treatment has been studied in a number of vulnerable groups. This information is presented here so that clinicians may have a better idea as to which groups are more likely to require a competency assessment. The present review is directed towards psychiatric and neurological conditions, although it should be noted that childhood and adolescent competency has been debated in different legal jurisdictions. There is evidence to suggest that children, and particularly adolescents, should have a greater degree of involvement in the treatment decision process. In one

---

2 In Ontario, the current legislation pertains to “capacity”. Hence, judgements are made in regards to capacity, not competency, which is discrepant with the definitions used in the present review.
landmark study, Weithorn and Campbell (1982) found that, relative to adults, 14-year-olds showed similar levels of capacity on four of the legal standards proposed by Roth et al. in 1977 (evidencing a choice, making a reasonable choice, rational reasons for a choice, and the ability to understand the facts relevant to a decision). Even 9-year-olds were able to evidence a choice and choose a reasonable treatment, although they were not as capable in regards to the standards of rational reasoning or understanding.

Low levels of capacity have been reported for other groups such as people with learning disability, geriatric psychiatry inpatient admissions, and medical inpatients (Arscott, Dagnan, & Kroese, 1999; Etchells et al., 1997; Mukherjee & Shah, 2001). Fitten, Lusky, and Hamann (1990) reported that the majority of nursing home residents had intermediate to profound impairments in capacity. Barton et al. (1996), using the HCAT, found 45% of nursing home admissions to score in the incompetent range. Given that normal older adults generally do not have impaired capacity, one would suspect that nursing home residents had poor capacity as a result of dementia, stroke, or other serious medical problems. However, the aforementioned studies did not specifically investigate the relationship between different diagnoses and capacity.

4.1. Alzheimer's disease

Several studies have shown that patients with mild and moderate Alzheimer’s disease have impaired decision-making capacity relative to normal older control groups, especially for more stringent abilities such as rational reasoning and understanding (Marson, Chatterjee, Ingram, & Harrell, 1996; Marson, Earnst, Jamil, Bartolucci, & Harrell, 2000; Marson, Hawkins, McInturff, & Harrell, 1997; Marson, Ingram, Cody, & Harrell, 1995; Kim et al., 2001). In these studies, normal older adults have reliably shown high capacity for all five legal standards, including those that are more difficult. Of concern is that, depending on the criteria used, very few patients with AD will have the capacity for consent. Even patients with mild forms of the disease will have serious problems with the more stringent components. When Marson et al. (1995) set cutoffs for each standard, they found that no patient with mild or moderate AD was competent on understanding, compared to 100% competency for normal older control subjects. Competency for patients with AD improved as the stringency of the standards decreased (e.g., 67% of patients with mild AD and 36% with moderate AD were competent on the appreciation standard).

In a subsequent study, Marson et al. (2000) showed that physicians, aided by the CCTI, rated 48% of AD patients to be competent (although only 34% were thought to possess adequate understanding). Kim et al. (2001) used experts to establish cutoff scores for competency on the MacCAT-CR and found that 62% of patients with AD were not competent for consenting to research. Cutoff scores based on normative data yielded an even higher percentage, with 84% of participants being non-competent.

4.2. Parkinson's disease

Dymek et al. (2001) studied capacity in Parkinson’s disease (PD) using the CCTI because it had proven validity with dementia of the Alzheimer’s type. Compared to a normal control group of older adults, patients with PD performed significantly worse on all legal standards, except for the controversial and biased standard of making a “reasonable choice.” As the legal standards increased in complexity, a clear pattern emerged of increasing incapacity.
4.3. Schizophrenia

In the MacArthur study, Grisso and Appelbaum (1995) compared schizophrenic, depressed, and angina inpatient groups to each other and matched community samples. Understanding (UTD) was significantly impaired in the schizophrenic group compared to the other hospitalized groups and the matched community sample. The performance of the schizophrenic patients on the UTD was significantly related to BPRS total scores (overall severity of symptoms) and BPRS factor III (thought disorganization). Appreciation was also impaired to the extent that one-third of the schizophrenia group scored poorly on the NOD subscore, indicating a failure to acknowledge their disorder. Other researchers have found poor insight to be a common feature of schizophrenia (Amador et al., 1993, 1994; Birchwood et al., 1994; Carroll et al., 1998; Cuffel, Alford, Fischer, & Owen, 1996; Debowska, Grzywa, & Kucharska-Pietura, 1998; David, Buchanan, Reed, & Almeida, 1992; Kemp & David, 1996; McEvoy et al., 1989; Michalakeas et al., 1994; Schwartz, 1998; Smith, Hull, & Santos, 1998; Smith et al., 1999; Swanson et al., 1995). Interestingly, on the NOT subtest, only 13% of the schizophrenia group did not acknowledge the value of treatment. The schizophrenia group performed worse on the TRAT-2 (reasoning) than any other diagnostic group and was significantly impaired in this regard relative to a matched community group on six of the eight cognitive functions related to reasoning (Grisso & Appelbaum, 1995).

When cutoffs were created for each of the measures to yield a dichotomous outcome for competence, patients with schizophrenia performed adequately on at least one measure. However, only 48% showed adequate capacity on all three measures, compared to 76% of the depressed group, 87.8% of the angina group, and 96% of the community group. The implications of these findings were that, on any given measure, only a minority of patients with schizophrenia showed impaired capacity. This minority tended to have more psychiatric symptoms, particularly thought disorder. The authors concluded that the “justification for a blanket denial of the right to consent to or refuse treatment for persons hospitalized because of mental illness cannot be based on the assumption that they uniformly lack decision-making capacity” (p. 171).

Grisso et al. (1997) used the MacCAT-T in a sample of 40 hospitalized schizophrenic patients and found them to perform significantly worse than a matched community control group on the understanding and reasoning abilities. Although the appreciation subtest did not apply to the control group, within the schizophrenia group, adequate ratings for appreciation were obtained by 83% of the sample and low ratings by only 10%. Moser et al. (2002) found that a schizophrenia group obtained significantly worse ratings than an HIV-positive group for the understanding and appreciation subtests of the MacCAT-CR, but not reasoning. Aside from reasoning, symptom severity and cognitive dysfunction significantly predicted understanding and appreciation. These investigators concluded that patients with schizophrenia as a whole displayed adequate decisional capacity using the MacCAT-CR and 80% had adequate understanding using the evaluation to sign consent (ESC).

Carpenter et al. (2000) presented a hypothetical research study to a group of 30 patients with schizophrenia and found that they performed worse on the MacCAT-CR relative to a normal control group. Those participants who had displayed any impairment on the MacCAT-CR understanding scale were given a 1-week educational program detailing the hypothetical study. They subsequently displayed such an improvement in their scores, which there was no longer a significant difference between the schizophrenia and control groups. The authors concluded that most people with schizophrenia will be capable of providing consent but a single session for the disclosure of a study may not be sufficient.
Dunn et al. (2002) examined the effect of enhanced consent procedures on comprehension in a sample of middle-aged and older adults with psychosis. One group of participants received a regular consent form, which was read aloud to them. Another group received the enhanced consent, which consisted of a Power Point presentation containing the same information as the regular consent form. Remarkably, the enhanced consent procedure was so effective that there was no difference in comprehension between the group that received this treatment and a comparison group of normal control subjects that were given the regular consent procedure. The EC procedure was repeated on several trials and it is noteworthy that, after three trials, virtually all of the patients had perfect comprehension. This points to the value of repeated teaching trials with this population.

To summarize, it would appear that only a minority of patients will perform poorly on any one component of competency and one should not assume that all patients with schizophrenia will be incompetent to provide consent. However, schizophrenia does reduce capacity relative to other psychiatric disorders and medical problems. Only half of schizophrenic patients can be expected to perform adequately on all four abilities. This does not mean that they will be unable to give consent. With the MacCAT-T or MacCAT-CR, investigators can pinpoint specific areas of incapacity and further education should remedy any impairments.

4.4. Depression

Grisso and Appelbaum (1995) found that for understanding (UTD), a hospitalized depressed group performed better than a schizophrenia group and no worse than a non-psychiatric group. Depressed subjects were not significantly different than the non-psychiatric group in terms of acknowledging that they had a disorder (NOD). This is consistent with several studies, which have shown depressed subjects to have high levels of insight (Amador et al., 1994; Michalakeas et al., 1994; Peralta & Cuesta, 1998; Pini, Cassano, Dell’Osso, & Amador, 2001; Sturman & Sproule, 2003). Despite the fact that the depressed patients were aware of their mental disorder, 14% did not acknowledge the potential for treatment (NOT). Depressed subjects thought more rationally (TRAT) about their treatment options than the schizophrenic group, but less so than the angina group or a matched control group. At least 90% of depressed subjects were capable on each measure and 75% were capable across all measures.

In 1999, Appelbaum et al. examined the competency of a group of depressed female outpatients. The majority of depressed subjects scored at a high level for each area of competency on the MacCAT-CR and their performance was not significantly related to depressive symptomatology. Therefore, depressed subjects should have little trouble providing informed consent if they are outpatients. Hospitalized subjects with more severe symptomatology will generally have high capacity but some may have difficulty thinking rationally about treatment.

4.5. Comparisons between groups

Vollmann, Bauer, Danker-Hopfe, and Helmchen (2003) have compared all of the aforementioned groups in a German sample. They tested the MacCAT-T in 31 patients with dementia, 43 patients with schizophrenia, and 35 patients with depression. Impairment for each component was defined with thresholds taken from the original MacArthur study (Grisso & Appelbaum, 1995). For the understanding component, patients with dementia had the most impairment (64.5%) followed by the schizophrenia (27.9%) and depressed (17.1%) groups. For reasoning, the dementia and schizophrenia groups had
similar levels of impairment (51.6% and 46.5%, respectively), whereas the depressed group was largely competent, with only 8.6% of patients showing impairment. None of the depressed patients had any impairment on the appreciation of disorder task compared to 22.6% of the patients with dementia and 16.3% with schizophrenia. Finally, only one individual from the depressed group and three from the schizophrenia group failed to appreciate possible treatment benefits. In this respect, the dementia group was markedly different with 32.3% displaying impairment.

In sum, the dementia group had high levels of impairment across the different abilities, particularly for understanding and reasoning where the majority could be said to lack competence. The schizophrenia group had similar problems with reasoning, but otherwise, the majority of patients were competent on the other components. And the depressed group showed very few problems in their capacity to consent for any one component.

5. Conclusions

Ultimately, whether or not a person is deemed to be capable of consent will depend on the legal customs and/or laws in a particular state, province, or country that relate to competence. Insofar as a standardized instrument separately assesses each component of capacity, it will remain useful under these circumstances. Also, a more inclusive measurement of capacity is preferable under most circumstances as competence judgements often incorporate information pertaining to unprecedented standards (Grisso & Appelbaum, 1995). Many of the available instruments measure all of the abilities relevant to competency, including the MacCAT-T, MacCAT-CR, CAT, SICIATRI, CCTI, and CIS. Of these instruments, the MacCAT-T and MacCAT-CR have been tested in more diagnostic categories than any others, and can be said to be the gold standards. Despite the fact that they have not been tested in large samples, these measures have repeatedly proven to be useful in detecting specific areas of incapacity. Nevertheless, some measures may be preferred over the MacCAT instruments for particular populations. For instance, the CCTI has been tested extensively in patients with dementia, especially Alzheimer’s disease. Likewise, the CAT may be a useful instrument for primary care patients. The CSA, as a measure of appreciation, is useful in conjunction with other measures in a competency assessment because, unlike other instruments, it does not confound appreciation with insight. Also, depending on the legal definition of competency in a particular jurisdiction, and the time constraints on researchers or clinicians, quick tests such as the HCAT may be desirable. However, researchers and clinicians alike should be aware that the majority of instruments, aside from the MacCAT-T, MacCAT-CR, and CCTI, require further testing.

Certain groups of individuals will be more likely to have impairments in capacity. The mentally ill will generally require more attention in this respect but should not be presumed to be incompetent. Furthermore, as Grisso and Appelbaum (1991) have pointed out, capacity will vary a great deal within and between diagnoses. For instance, depressed subjects can be expected to have sufficient capacity for consent relative to other psychiatric groups, but some may have greater difficulties if they are hospitalized or psychotic. People with schizophrenia are more likely to be incapable of consent but, even in this case, about half of all patients will be competent in all possible areas. And for those who are not competent on one or more standards, education can be expected to lead to improvements. Finally, the capacity to give informed consent should be investigated in other patient groups such as bipolar disorder and substance abuse.
Acknowledgements

Gordon Duval and Carole Cohen offered many valuable suggestions for this paper and Harvey Brooker, Beth Sproule, Philip Hebert, Vesna Sturman and Myriam Mongrain were generous with their time and help.

References


