

## The inter-rater reliability of mental capacity assessments

Vanessa Raymont<sup>a</sup>, Alec Buchanan<sup>b</sup>, Anthony S. David<sup>a</sup>, Peter Hayward<sup>a</sup>,  
Simon Wessely<sup>a</sup>, Matthew Hotopf<sup>a,\*</sup>

<sup>a</sup> Institute of Psychiatry, King's College, London

<sup>b</sup> Yale University, New Haven, USA

Received 10 December 2004; received in revised form 15 September 2005; accepted 25 September 2005

---

### Abstract

*Background:* Assessing mental capacity involves complex judgements, and there is little available information on inter-rater reliability of capacity assessments. Assessment tools have been devised in order to offer guidelines. We aimed to assess the inter-rater reliability of judgements made by a panel of experts judging the same interview transcripts where mental capacity had been assessed.

*Method:* We performed a cross sectional study of consecutive acute general medical inpatients in a teaching hospital. Patients had a clinical interview and were assessed using the MacArthur Competence Assessment Tool for Treatment (MacCAT-T) and Thinking Rationally About Treatment (TRAT), two capacity assessment interviews. The assessment was audiotaped and transcribed. The raters were asked to judge whether they thought that the patient had mental capacity based on the transcript. We then divided participants into three groups — those in whom there was unanimous agreement that they had capacity; those in whom there was disagreement; and those in whom there was unanimous agreement that they lacked capacity.

*Results:* We interviewed 40 patients. We found a high level of agreement between raters' assessments (mean kappa=0.76). Those thought unanimously to have capacity were more cognitively intact, more likely to be living independently and performed consistently better on all subtests of the two capacity tools, compared with those who were unanimously thought not to have capacity. The group in whom there was disagreement fell in between.

*Conclusions:* This study indicates that clinicians can rate mental capacity with a good level of consistency.

© 2006 Elsevier Inc. All rights reserved.

*Keywords:* Mental capacity; Competence; Assessment; Inter-rater reliability

---

### 1. Introduction

Mental capacity is variously defined, but in recent English legislation ([Mental Capacity Act, 2005](#)) a patient lacks capacity if there is “an impairment of or disturbance in the functioning of brain or mind” and this causes difficulty in decision making because the individual is: (a) unable to understand the information relevant to the decision; or (b) unable to retain the information relevant to the decision; or (c) unable to use or weigh that information as part of the process of making the decision, or (d) unable to communicate the decision. Capacity is vital in ensuring a person can

---

\* Corresponding author. Department of Psychological Medicine, Institute of Psychiatry, Weston Education Centre, Cutcombe Rd, London SE5 9RJ, United Kingdom. Tel.: +44 207 848 0778; fax: +44 207 848 5408.

E-mail address: [m.hotopf@iop.kcl.ac.uk](mailto:m.hotopf@iop.kcl.ac.uk) (M. Hotopf).

exercise his or her autonomy in making choices, including medical or psychiatric treatment. In England and Wales, current clinical practice utilizes a functional approach to capacity assessment, based on legal definitions of discrete but complementary components of competence (Lord Chancellor, 1999). A ‘sliding scale’ of incapacity is often employed (Wong, Clare, Gunn, & Holland, 1999; Buchanan, 2004), with the threshold for incapacity reflecting the complexity, likely outcome and significance of the decision. Mental capacity may fluctuate with time, and the same individual may have capacity for some decisions, whilst lacking capacity for others. All these considerations mean that judging mental capacity is a complex process, which clinicians may find difficult to do (Raymont, 2002).

Researchers require reliable and valid measures to study clinical problems such as mental capacity. Current capacity assessment measures include extensive research-based instruments (Grisso & Appelbaum, 1995), and briefer tools designed to be used by clinicians. The latter include semi-structured interviews largely based on legal criteria (Draper & Dawson, 1990), recognition tests (Janofsky, McCarthy, & Folstein, 1992), and patient-centred clinical vignettes (Fazel, Hope, & Jacoby, 1999). One well-established tool is the MacArthur Competence Assessment Tool for Treatment (MacCAT-T) (Grisso, Appelbaum, & Hill-Fotouhi, 1997). This is a semi-structured interview, which is designed to examine four components of capacity in a clinical setting: (1) understanding of the disorder and its treatment, including associated benefits/risks; (2) appreciation of the disorder and its treatment (requiring insight into how these will affect the patient individually); (3) reasoning, which examines how and why the decision was made, and the potential to compare consequences; and (4) the ability to express a choice. The MacCAT-T rating allows the detection of inadequacies in any of these four areas, but the authors emphasize that it does not give an overall rating and should always be used to make a judgement of capacity in conjunction with a clinical assessment. The tool has been used in a variety of diagnostic groups, including individuals with affective disorders and psychotic illness (Grisso et al., 1997; Carpenter et al., 2000; Vollmann, Bauer, Danker-Hopfe, & Helmchen, 2003; Palmer, Dunn, Appelbaum, & Jeste, 2004). The MacCAT-T was developed from the original MacArthur Treatment Competence Research Instruments for assessing capacity to consent to treatment; Understanding Treatment Disclosures (UTD), Perceptions of Disorder (POD) and Thinking Rationally About Treatment (TRAT) (Appelbaum & Grisso, 1995). These assess the person’s understanding of information about treatment as would be required to a consent procedure, their appreciation of significance of information about disorder or treatment when applied to their own circumstances, and the quality of their cognitive functions employed in making a treatment decision respectively. They have been validated in a sample of hospitalized patients with schizophrenia or schizoaffective disorder, major depression and ischaemic heart disease (Grisso & Appelbaum, 1995).

These assessments do not generate a categorical result (i.e. “has capacity” vs “does not have capacity”) and the authors suggest they are used alongside a clinical interview, in order to make such a judgement. We wanted to determine whether experienced clinicians reading transcripts of interviews based on MacCAT-T, could agree on categorical decisions about whether the patient had capacity to make a current treatment decision. Although a number of studies (e.g. Wong, Clare, Holland, Watson, & Gunn, 2000; Etchells et al., 1999 and Roth et al., 1982) have reported inter-rater reliabilities for mental capacity assessments based on other interviews, we are not aware of this having been reported for the MacCAT-T. We wanted to measure this in a mixed group of acutely ill general hospital inpatients, in order to reflect normal clinical practice in a population likely to have high rates of impairments of decision making ability.

## 2. Method

We obtained local research ethics committee approval for the study. We assessed the capacity of a mixed sample of adults (aged 18 years upwards) on two acute general medical wards in a London teaching hospital. The wards admitted all adults with medical conditions over the age of 18 years of age. A random sample of all eligible patients was invited to take part in the interview. A psychiatrist (VR) established if participants in this group had a current diagnosis and active treatment or planned treatment. Those meeting these criteria were then interviewed between 48 h and 7 days after admission. Those judged by their clinical team to be too distressed to take part were excluded. We did not approach those obviously lacking capacity because of lack of consciousness or because they could not communicate. We also excluded any patient with extreme cognitive impairment (assessed by inability to give his or her date of birth) and those unable to communicate adequately in English. The present study describes a subsample of the main study which aimed to determine prevalence of and risk factors for mental incapacity (Raymont et al., 2004).

The MacArthur Competence Assessment Tool-Treatment (MacCAT-T) (Grisso et al., 1997) was administered by a specialist registrar in psychiatry (VR). For the purposes of these assessments we defined the current treatment as the

most recently instituted intervention used to treat the main reason for admission. We made one alteration to the MacCAT-T as it was not considered appropriate for the research team to provide diagnostic information. We therefore dropped the understanding of disorder component of the interview, and the understanding dimension is on a 0–4 instead of the usual 0–6 scale. VR also administered vignettes based on the Thinking Rationally About Treatment assessment (TRAT — (Grisso & Appelbaum, 1992)). We used the original vignette based on ischaemic heart disease and devised four others with a similar format, which gave greater clinical relevance to our sample as they were based around common physical illnesses. The MacCAT-T and TRAT interviews were audiotaped and transcribed. All participants were administered the Mini Mental State Examination (MMSE) (Folstein, Folstein, & McHugh, 1975).

Performance on the MacCAT-T and the vignette was assessed by five raters- four consultant psychiatrists (a neuropsychiatrist (AD), two consult-liaison psychiatrists (MH and SW) and a forensic psychiatrist (AB)) and one consultant clinical psychologist (PH), all with extensive experience of mental capacity assessments. The raters were asked to make a judgement as to whether an individual had, or did not have mental capacity to make the treatment decision, based on the legal definition which had been proposed in English Law (Lord Chancellor, 1999). Raters did not have the option of rating a transcript as “don’t know”. The raters performed their ratings in isolation (i.e. were not able to confer) and were blind to the patient’s performance on the MMSE and sociodemographic variables (e.g. age, sex and social class).

We collected demographic and clinical information, including whether the patient needed any assistance in activities of daily living, or was able to live independently. We determined the nature of the current illness, and whether this was a recurrent or new problem. We determined the total number of diagnoses and medications.

### 2.1. Statistical analysis

The ratings were compared using the kappa test, which gives an indication of the degree of agreement above that expected by chance. Participants were then categorised according to whether the raters judged them unanimously to have capacity or not to have capacity, or whether there was disagreement between raters. These groups were compared on clinical and demographic characteristics and performance of the individual subtests of the MacCAT-T and TRAT, rated by the interviewer. For each MacCAT-T and TRAT subscale we computed the proportion of individuals who scored greater than or equal to 50% marks. Owing to small cell sizes we used Fisher’s exact test for categorical variables. Non-parametric testing (Kruskal Wallace test) was used for continuous variables.

## 3. Results

A total of 159 patients were interviewed in our main study. This study describes the inter-rater reliability measures for a subsample of the first 40 consecutive interviews. This group did not differ significantly in terms of their demographic or clinical characteristics from the main sample (age;  $t=0.8$ ,  $p=0.4$ ; gender;  $\chi^2=9.8$ ,  $p=0.2$ ). The characteristics of non-responders is described in greater depth in the companion paper (Raymont et al., 2004). The mean age of our sample was 64.2 years, 47.5% were female, and the three most common diagnostic groups were cardiovascular (25.0%) and respiratory disease (25.0%) and current infection (17.5%).

### 3.1. Agreement between raters

The kappa scores for individual clinicians based on their rating of interview transcripts are shown in Table 1. We found a strong level of agreement between individual raters’ assessments in our subsample (mean kappa=0.76).

Table 1  
Agreement between raters

Rater no.	2	3	4	5
1	.82	.83	.61	.94
2	–	.88	.78	.77
3		–	.66	.77
4			–	.57

Values are kappas.

Mean kappa=0.76.

Table 2  
Characteristics of participants according to capacity assessment

	Has capacity unanimously	Disagreement on capacity	Lacks capacity unanimously	Significance ( <i>p</i> )
<i>N</i>	24 (60.0%)	9 (22.5%)	7 (17.5%)	
Mean age (median)	62.5	70.0	80.0	0.006
Gender (% males)	50.0	66.7	42.9	0.7
Ethnicity (% white)	91.7	100.0	71.4	0.16
% Living independently (including with family input)	87.5	77.8	42.8	0.05
Years of education (mean)	11.0	9.7	10.3	0.02
% with recurrent diagnosis, same presentation	62.5	22.2	14.3	0.03
% clinical team thought had capacity	100	88.9	71.4	0.03
% nearest relative thought had capacity	100	75.0	75.0	0.004
MMSE (median)	28.5	26.0	16.0	<i>p</i> <0.001

### 3.2. Characteristics of the sample

Table 2 compares the three groups on demographic and clinical variables. We found that groups did not differ in terms of gender or ethnic group. However those judged to lack capacity were older, less likely to be living independently, less likely to be presenting with a recurrent illness, but to score significantly lower on the MMSE. Individuals we thought lacked capacity were more likely to be thought to lack capacity by their nearest relative or clinician. The group where there was disagreement on ratings of capacity fell between the other two groups on each of these variables. It is noteworthy that even in the group who we unanimously agreed lacked capacity, the relatives and clinicians in the majority of cases thought that the patient had capacity.

Table 3  
Performances on MacCAT-T subtests

	Has capacity unanimously	Disagreement on capacity	Lacks capacity unanimously	Significance ( <i>p</i> )
MacCAT-T understanding summary rating (%)				
4	75.0	22.2	0	
3	25.0	11.1	0	
2	0	55.6	28.6	
1	0	0	42.9	
0	0	11.1	28.6	
≥50% total score (%)	100	88.9	28.6	<i>p</i> <0.001
MacCAT-T appreciation summary rating (%)				
4	66.7	22.2	0	
3	25.0	33.3	14.3	
2	8.3	22.2	28.6	
1	0	22.2	14.3	
0	0	0	42.9	
≥50% total score	100	77.8	42.9	<i>p</i> <0.001
MacCAT-T reasoning summary rating (%)				
8	0	0	0	
6–7	33.3	11.1	0	
4–5	50.0	0	0	
2–3	16.7	55.6	28.6	
0–1	0	33.3	71.4	
≥50% total score	83.3	11.1	0	<i>p</i> <0.001
MacCAT-T expressing a choice				
2	90.9	66.7	28.6	<i>p</i> =0.002
1	9.1	11.1	57.1	
0	0	22.2	14.3	
% with ≥50% score	100	77.8	85.7	

Table 4  
Performances on TRAT subtests

	Have capacity unanimously	Disagreement on capacity	Lack capacity unanimously	Significance ( <i>p</i> )
Seeking information % with $\geq 50\%$ score	45.8	22.2	0	0.17
Vignette consequential thinking % with $\geq 50\%$ score	100	100	42.9	0.002
Vignette comparative thinking % with $\geq 50\%$ score	66.7	55.6	14.3	0.17
Vignette complex thinking % with $\geq 50\%$ score	91.7	88.9	42.9	0.05
Vignette generating consequences % with $\geq 50\%$ score	95.8	88.9	71.4	0.04

### 3.3. Performance on MacCAT-T and TRAT

In all subtests of the MacCAT-T those with capacity performed better than those lacking capacity and those in whom capacity was questionable (Table 3). This was especially so in the reasoning subscores. Those judged not to have capacity, or with doubtful capacity, produced lower scores on all the vignette items, when compared with those with capacity. However, the disparity was less striking, and only reached significant levels in the consequential thinking, complex thinking and generating consequences questions (which again represented the reasoning components of the test). However, all those lacking capacity failed to ask for any further information before making their treatment decision, which may represent a deficiency in initiation of decisions (so called ‘intentionality’ — (Grimes, McCullough, Kunik, Molinari, & Workman, 2000). A similar pattern was found for the TRAT (Table 4).

## 4. Discussion

This study indicates that using transcripts from a semi-structured interview assessing capacity, experienced psychiatrists and psychologists are able to make reasonably consistent judgements, despite a heterogenous clinical population with diverse medical disorders. This suggests that for further research interviews based on these measures can be used reliably to determine capacity in clinical populations.

It is reassuring that those judged by the panel to lack capacity are much more likely to have cognitive impairments on the MMSE; to be less likely to live independently and to perform consistently less well on all subtests of the measures, despite the panel being blind to this information. This suggests that the assessment has some degree of concurrent validity.

Finally it is also reassuring that the three groups we described (unanimously has capacity; disagreement on capacity; and unanimously lacks capacity) performed in a predictable way with a gradient between those unanimously judged to have capacity who performed best on each subtest of the capacity measures and on the MMSE, with the disagreement group performing less well, and the unanimously lacking capacity group performing least well. This suggests that the clinicians in the panel made judgements that a patient lacked capacity at slightly different thresholds, but that they probably used similar information in each case to make the judgement.

The main methodological weakness of this study is that the panel made decisions on capacity based on a single transcribed interview. This is different from the situation in clinical practice, where different clinicians would interview the patient at different points in time. Whilst this was not possible in the present study, we plan to perform a further study to test the reliability of assessments of capacity based on two separate interviews performed by different clinicians.

## Acknowledgement

This research was funded by a grant from the Wellcome Trust.

## References

- Appelbaum, P. S., & Grisso, T. (1995). The MacArthur Treatment Competence Study. I: Mental illness and competence to consent to treatment. *Law and Human Behavior*, 19, 105–126.
- Buchanan, A. (2004). Mental capacity, legal competence and consent to treatment. *Journal of the Royal Society of Medicine*, 97, 415–420.
- Carpenter, W. T., Gold, J. M., Lahti, A. C., Queern, C. A., Conley, R. R., Bartko, J. J., et al. (2000). Decisional capacity for informed consent in schizophrenia research. *Archives of General Psychiatry*, 57, 533–538.

- Draper, R. J., & Dawson, D. (1990). Competence to consent to treatment: A guide for the psychiatrist. *Canadian Journal of Psychiatry*, 35, 285–288.
- Etchells, E., Darzins, P., Silberfeld, M., Singer, P. A., McKenny, J., Naglie, G., et al. (1999, January). Assessment of patient capacity to consent to treatment. *Journal of General Internal Medicine*, 14(1), 27–34.
- Fazel, S., Hope, T., & Jacoby, R. (1999). Assessment of competence to complete advance directives: Validation of a patient centred approach. *British Medical Journal*, 318, 493–497.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.
- Grimes, A. L., McCullough, L. B., Kunik, M. E., Molinari, V., & Workman, R. H. (2000). Informed consent and neuroanatomic correlates of intentionality and voluntariness among psychiatric patients. *Psychiatric Services*, 51, 1561–1567.
- Grisso, T., & Appelbaum, P. S. (1992). *Manual for rational thinking about treatment*. Massachusetts: Massachusetts Medical School.
- Grisso, T., & Appelbaum, P. S. (1995). The MacArthur treatment competence study III: Abilities of patients to consent to psychiatric and medical treatments. *Law and Human Behavior*, 19, 149–174.
- Grisso, T., Appelbaum, P. S., & Hill-Fotouhi, C. (1997). The MacCAT-T: A clinical tool to assess patients’ capacities to make treatment decisions. *Psychiatric Services*, 48, 1415–1419.
- Janofsky, J. S., McCarthy, R. J., & Folstein, M. F. (1992). The Hopkins Competency Assessment Test: A brief method for evaluating patients’ capacity to give informed consent. *Hospital and Community Psychiatry*, 43, 132–136.
- Lord Chancellor (1999). *Making decisions: The government’s proposals for making decisions on behalf of mentally incapacitated adults*. London: The Stationary Office.
- Mental Capacity Act. (2005).
- Palmer, B. W., Dunn, L. B., Appelbaum, P. S., & Jeste, D. V. (2004). Correlates of treatment-related decision-making capacity among middle-aged older patients with schizophrenia. *Archives of General Psychiatry*, 61, 230–236.
- Raymont, V. (2002). ‘Not in perfect mind’ — the complexity of clinical capacity assessment. *Psychiatric Bulletin*, 26, 201–204.
- Raymont, V., Bingley, W., Buchanan, A., David, A. S., Hayward, P., Wessely, S., et al. (2004). The prevalence and associations of mental incapacity in medical in-patients. *Lancet*, 364, 1421–1427.
- Roth, L. H., Lidz, C. W., Meisel, A., Soloff, P. H., Kaufman, K., Spiker, D. G., et al. (1982). Competency to decide about treatment or research. *International Journal of Law and Psychiatry*, 5, 29–50.
- Vollmann, J., Bauer, A., Danker-Hopfe, H., & Helmchen, H. (2003). Competence of mentally ill patients: A comparative empirical study. *Psychological Medicine*, 33, 1463–1471.
- Wong, J. G., Clare, I. C. H., Gunn, M., & Holland, A. J. (1999). Capacity to make healthcare decisions: Its importance in clinical practice. *Psychological Medicine*, 29, 437–446.
- Wong, J. G., Clare, I. C. H., Holland, A. J., Watson, P. C., & Gunn, M. (2000, March). The capacity of people with a ‘mental disability’ to make a health care decision. *Psychological Medicine*, 30(2), 295–306.