Law and Human Behavior, Vol. 19, No. 2, 1995

The MacArthur Treatment Competence Study. II

Measures of Abilities Related to Competence to Consent to Treatment*

**Thomas Grisso,† Paul S. Appelbaum,†
Edward P. Mulvey,‡ and Kenneth Fletcher†**

This article reports the development and psychometric properties of three standardized and objectively scored measures, the MacArthur Treatment Competence Research Instruments. They were designed to assess abilities related conceptually to four legal standards for competence to consent to treatment: understanding, appreciation, rational manipulation (reasoning), and expressing a choice. Scoring reliability, internal consistency, intertest correlations, and test–retest correlations were examined with data from samples of hospitalized patients with schizophrenia, major depression, and ischemic heart disease, as well as matched non-ill community samples. The results indicate very good interscorer reliability and provide guidance for the use of the instruments and interpretation of their results in future research on patients' decisional abilities in treatment contexts.

There are two important reasons to develop measures of abilities that are related conceptually to competence to consent to treatment among persons with mental illness (Appelbaum & Grisso, 1995). First, during the past two decades, developments in law (e.g., *Rivers v. Katz,* 1986; *Rogers v. Commissioner,* 1983) and ethics

* This research was supported by the John D. and Catherine T. MacArthur Foundation Research Network on Mental Health and the Law. Other members of the Network are acknowledged for their conceptual guidance during the process of the study: Shirley Abrahamson, Richard Bonnie, Pamela Hyde, John Monahan, Stephen Morse, Loren Roth, Paul Slovic, Henry Steadman, and David Wexler, as well as Steven K. Hoge. Acknowledged for their role in data collection management are Deidre Klassen and Cynthia Wickless. Requests for manuals for the MacArthur Treatment Competence Research Instruments and other correspondence may be addressed to either author at Department of Psychiatry, University of Massachusetts Medical School, Worcester, MA 01655.

† University of Massachusetts Medical School.

‡ Western Psychiatric Institute and Clinic, University of Pittsburgh.

(e.g., President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research, 1982) raised policy-related questions that require answers based on controlled research on patients' abilities to make treatment decisions. To provide reliable information related to policy, this research requires standardized instruments for assessing the relevant abilities.

Second, these same developments have increased the need to evaluate patients' capacities to participate in treatment decisions on a case-by-case basis (Cutter & Shelp, 1991). Legal protection of the right of competent patients with mental illness to refuse treatment produces the need for reliable and valid clinical assessment of patients' decision-making abilities in those cases in which a patient's competence is questioned.

This article describes the development of three instruments (the MacArthur Treatment Competence Research Instruments) designed for research purposes to assess the abilities of persons with mental illness to participate in their own treatment decisions. The first article in this series (Appelbaum & Grisso, 1995) reviewed past research in this area and described the design of the study for which the instruments were created, and the third article in the series presents the results (Grisso & Appelbaum, 1995).

The instruments assess abilities related conceptually to four legal standards used to determine the competence of patients to consent to treatment (Appelbaum & Grisso, 1988, 1995): understanding, appreciation, rational manipulation (reasoning), and expressing a choice. (See Grisso, 1986, for a more detailed description of the approach used here for the development of instruments to assess psychological abilities related to legal standards.) We operationalized the relationships between the legal standards and relevant psychological functions in the following ways.

For abilities related to *understanding*, the type of information that a person must be capable of comprehending was identified for our purposes by the requirements of informed consent: (a) the nature of the patient's disorder, (b) the nature of the treatment that is being recommended, (c) the probable benefits of the treatment, (d) the probable risks and discomforts of the treatment, and (e) any alternative treatments and their relevant benefits and risks (Appelbaum, Lidz, & Meisel, 1987). Therefore, we operationalized *understanding*, a functional ability related to competence to consent to treatment, as a person's ability to demonstrate comprehension of such information by paraphrasing or recognizing items of information (related to one's own mental disorder) after they are presented in an informed consent disclosure.

Concerning *appreciation*, the legal concept refers to patients' recognition that information given to them about their disorder and potential treatment is significant for and applicable to their own circumstances. Not all disagreements with diagnosis or treatment, however, raise questions of legal incompetence. Those based on rigidly held beliefs that involve distortion or denial as mechanisms of defense or symptoms of psychopathology most strongly suggest incompetence (Appelbaum & Grisso, 1992). Therefore, we operationalized *appreciation*, as it relates to competence to consent to treatment, as acknowledgment of illness and the potential value of treatment, or acknowledgment of these things after illogical

premises underlying initial nonacknowledgment were challenged. (Note that appreciation focuses on patients' acknowledgment of the *potential* value of treatment; whether they would actually accept the treatment is not relevant.)

Courts' descriptions of the legal standard for *rational manipulation* (reasoning) refer to the quality of an individual's cognitive abilities that are employed in the process of using information to arrive at a decision (Appelbaum & Grisso, 1988, 1995; Grisso, 1986; Grisso & Appelbaum, 1993; President's Commission, 1982; Roth, Meisel, & Lidz, 1977; Tepper & Elwork, 1984). This standard refers to the quality of the individual's mental operations while deciding, not the quality, rationality, or reasonableness of the person's actual choice. The focus is on whether the decision is the product of a logical and adequate *reasoning process*, regardless of the accuracy or reasonableness of the premises employed in the reasoning. As we describe later, decision-making models in psychology have identified a number of abilities that are relevant for problem-solving situations in general. Therefore, we operationalized *reasoning*, a functional ability related to competence to consent to treatment, as one's demonstration of several of these problem-solving abilities when faced with a decision about treatment for a disorder.

The ability to *communicate a choice* has been considered by courts when individuals, because of illness, have been unable to reach a decision or to maintain a stable choice. We operationalized this ability as one's selection of a treatment option in a decision-making task.

METHOD

Participants

A companion article (Grisso & Appelbaum, 1995) describes in detail the characteristics of the samples that contributed the data on which reliability and structural analyses of the MacArthur Civil Competence Research Instruments were performed. The two samples with mental illness consisted of 75 patients diagnosed with schizophrenia or schizoaffective disorder and 92 patients with major depressive disorder, all of whom had been admitted recently to psychiatric units at the time that they were administered the measures. A third sample consisted of 82 patients with no mental disorder who had been admitted to medical hospitals for evaluation or treatment of ischemic heart disease (angina). These three samples were referred to collectively as the *hospitalized groups*.

Three non-ill and nonhospitalized samples (the *community groups*) consisted of individuals in communities in the catchment areas of the hospitals from which the samples with medical illness and mental illness were drawn. Each community group represented a comparison sample for one of the hospitalized samples, matched person-for-person on age, gender, race, and socioeconomic status. Sample sizes were identical to the ill groups with mental illness and medical illness for which they provided control samples.

Each sample contained participants from at least two sites among three re-

search sites participating in the study. The Worcester, MA, site provided 73% of the angina group and 25% of the schizophrenia group; Kansas City, MO, provided 75% of the schizophrenia group and 29% of the depression group; and Pittsburgh, PA, provided 71% of the depression group and 27% of the angina group. Each site also provided the community participants that contributed in similar proportions to the control groups.

Measures

All participants were administered the Beck Depression Inventory (BDI: Beck, 1978), and three subtests of the Wechsler Adult Intelligence Scale–Revised (Vocabulary, Similarities, Digit Span) with which verbal IQ was prorated to serve as an index of verbal intellectual functioning. In addition, hospitalized participants received the Brief Psychiatric Rating Scale (BPRS: Overall, 1988).[1]

Three instruments were developed as measures of abilities that were conceptually related to four legal standards for competence to consent to treatment (see companion article, Appelbaum & Grisso, 1995, for descriptions of the legal standards). Manuals for the following three instruments are available from the authors on request.

Understanding Treatment Disclosures (UTD)

The UTD provides research data relevant for psychological inferences that are made in legal inquiries applying the understanding standard for competence to consent to treatment (Appelbaum & Grisso, 1988, 1995; Grisso & Appelbaum, 1991, 1992). It assesses a person's understanding of information similar to that which is disclosed to patients in an informed consent process.

Administration of the UTD involves the standardized presentation of five paragraphs of information corresponding to content required for informed consent to treatment (Appelbaum, Lidz, & Meisel, 1987). The information is called the *disclosure*, and each paragraph constitutes an *element* of the disclosure. The disclosure was designed to represent all five content elements for informed consent, but it was not meant to be a replica of the full disclosure that might be required in various clinical situations.

Three forms of the disclosure were developed for the present research, for use with each of the patient groups that were studied: schizophrenia, depression, and ischemic heart disease. This was in keeping with our principle that the content of the instruments should be meaningful to participants to eliminate the potentially confounding effects of using information not relevant to each patient's situation (Appelbaum & Grisso, 1995). The content of each disclosure form reflects its specific disorder and its treatment options. Each of the three forms has the same format (five paragraphs of two-to-five sentences each) and absolute length, and

[1] The version of the BPRS used here included the usual 18 items (Overall, 1988) plus one additional item (*elevated mood*). As Overall (1988) explained, this 19-item version was in use during the 1960s at the NIMH Psychopharmacology Service Center, although there is no publication reference for this version. Research results are very unlikely to differ for the 18-item vs. the 19-item version.

they are worded to meet a 7th to 9th grade reading ease criterion (Grunder, 1978). A printed copy of the disclosure is given to respondents to read while it is read aloud to them.

The UTD procedure provides for two types of disclosure: *uninterrupted* and *element* disclosure. Uninterrupted disclosure (UD) involves presentation of all five elements prior to assessment of the person's understanding. Performance in this task requires understanding and recalling information that is embedded in a larger message that includes many pieces of information. Element disclosure (ED), in contrast, involves the presentation of each element separately, with understanding assessed following each element. This creates less of a demand on storage and retrieval of specific information from a larger body of information.

Finally, the UTD provides for assessment of understanding using two modes of response. *Paraphrased recall* (PR) requires respondents to express their understanding of the disclosed information in their own words in response to standardized questions and inquiry probes. *Recognition* (RC) presents respondents with statements that either do or do not offer a message that was contained in the disclosure and asks them simply to identify whether it is the "same as" or "different from" the messages in the disclosure.

The research for which the UTD was developed employed three combinations of the above disclosure and response modes with each respondent, always in the same sequence. The first procedure used uninterrupted disclosure with paraphrased recall (hereinafter, Uninterrupted–Paraphrase). Then each element was presented separately, with paraphrased recall (Element–Paraphrase) and recognition (Element–Recognition) procedures both occurring at the end of an element's disclosure prior to moving to the next element. These three procedures are referred to as the UTD subtests. They require about 25–30 min for administration.

The UTD manual provides objective scoring criteria and examples for each subtest, with different scoring versions for each of the three forms of disclosure. Scores are received (2, 1, 0 credit) on each of the five elements in a subtest. Element scores are summed to produce a subtest score (0–10) for each UTD subtest separately.

Perceptions of Disorder (POD)

The POD provides research data relevant for psychological inferences that are made in legal inquiries applying the appreciation standard for competence to consent to treatment (Appelbaum & Grisso, 1988, 1992, 1995). It assesses patients' failure to appreciate the significance of information about disorder or treatment when applied to their own circumstances. The POD has two parts, measuring nonacknowledgment of one's disorder, and nonacknowledgment of the potential value of treatment even when successful treatment is likely. Clinical references to these beliefs frequently refer to them as forms of "denial," or "lack of insight," that reflect cognitive incapacity, delusional thinking, or defensive distortions of reality that protect patients psychologically from painful emotions or thoughts associated with their condition.

Patients may disavow (fail to acknowledge) a diagnosed disorder or the po-

tential value of treatment for reasons other than psychiatric disorder or psychological distortion. Disavowal of the potential value of treatment, for example, may occur because of religious and cultural beliefs (e.g., a belief in religious healing that conflicts with medical interventions) or for reasons that are based on experience (e.g., a past treatment history that includes a number of unsuccessful medication trials). In contrast, the POD was designed to identify disavowal that is more likely to be defensive or delusional in nature and which is asserted rigidly (therefore less adaptively) in the face of disconfirming information (Appelbaum & Grisso, 1992).

The POD offers a standardized interview procedure that includes nine stimulus questions. Three of these inquiries assess patients' degree of acknowledgment of their disorder, including (a) acknowledgment of their symptoms as provided to them from their hospital chart, (b) their beliefs about the severity of their symptoms, and (c) their acknowledgment of the formal diagnosis that is provided to them from their hospital chart. These items are grouped as the *Nonacknowledgment of Disorder* (NOD) subtest.

Another three questions pertain to patients' degree of acknowledgment that treatment might be of benefit to them, including acknowledgment (a) of the relevance of obtaining any treatment (generally) for one's condition, (b) of the potential benefit of a specific proposed treatment (medication), and (c) of the lesser likelihood of improvement without the treatment. These items comprise the *Nonacknowledgment of Treatment Potential* (NOT) subtest.

Three forms of the POD were developed for the present research, corresponding to patient groups that were studied: schizophrenia, depression, and ischemic heart disease. The content of the items in each form reflects its specific disorder and its treatment options. The POD procedure focuses respondents on their beliefs about their own psychiatric or medical condition. Therefore, information about specific symptoms and diagnosis are taken from respondents' hospital charts and inserted in the items at relevant points.

Each of the POD items begins by providing a piece of information to the respondent. The respondent then is asked to express his or her own belief concerning the validity of the information that was provided. The respondent assigns a value to this belief on a 6-point scale that is anchored with labels such as *definitely no* to *definitely yes*. The respondent then is asked to explain his or her belief. (See Table 1 for an example.) Beyond this similarity, procedures on the two subtests differ in the following ways.

On the Nonacknowledgment of Disorder subtest, respondents' belief ratings constitute the data for scoring. For the first and third items (acknowledgment of symptoms and of diagnosis), full nonacknowledgment (0 points) is given only if respondents select either of the two ratings indicating strong disagreement and partial nonacknowledgment (1 point) if they provide either of the two middle ratings on the 6-point scale. For the second item (referring to severity of symptoms), a method is used to assign a nonacknowledgment score to members of the mentally ill groups based on the correspondence between the respondent's rating and the respondent's Brief Psychiatric Rating Scale score (BPRS, a measure of symptom severity used in the present study). For example, a full nonacknowl-

Table 1. Example of POD (Appreciation) Item and Use of Questioning with a Hypothetical Situation

Item D: Acknowledgment of Potential for Treatment

Questioning: "Most people who have symptoms of a mental or emotional disorder like your doctor believes you have can be helped by treatment. The most common treatment is medication. Other treatments sometimes used for such disorders are having someone to talk to about problems, and participating in group therapy with other people with similar symptoms."

1. "Using the card again" [referring to 6-point scale: see accompanying text], "do you believe that you have the kind of condition for which some types of treatment might be helpful?"

2. "All right, you believe that . . ." [use verbal anchor for the number on the scale chosen by patient, to paraphrase the patient's expressed opinion]. "Can you explain that to me? What makes you believe that . . ." [again paraphrase as above]? *If answer on six-point continuum is any of the three levels expressing low value of treatment,* find appropriate hypothetical, present it, and obtain second opinion rating and explanation.

Example of Hypothetical

[For patient who believes that treatment will not work because he or she is "just too sick:"]

"Imagine that a doctor tells you that there is a treatment that has been shown in research to help 90% of people *with problems just as serious as yours.* Do you think this treatment might be of more benefit to you than getting no treatment at all?"

edgment score (0 points) on this item is obtained by a respondent who rates his or her symptoms as not severe (two lowest ratings on the 6-point scale) when the respondent's BPRS score is above 40. Angina patients were scored based on the assumption that ischemic heart disease severe enough to warrant hospitalization was of substantial severity. Respondents who were able to acknowledge at least partial awareness of this (four highest ratings on the 6-point scale) received a score of 2, whereas those who failed to acknowledge any degree of severity (lowest rating on the scale) received a score of 0 points.

For the Nonacknowledgment of Treatment Potential subtest, any rating on the acknowledgment side of the 6-point scale is scored 2—that is, full appreciation. A special procedure, however, is required when respondents provide ratings initially indicating any degree of nonacknowledgment. This is because individuals might fail to acknowledge the potential value of treatment for many reasons that would not meet our definition of failure of appreciation (as defined earlier in this report). For example, a person may believe in the nonefficacy of treatment simply because he or she has been disappointed by doctors' treatment efforts in the past.

Therefore, when any degree of nonacknowledgment is expressed on an item in this subtest, the examiner asks for respondents' reasoning for their ratings. They are then presented with a hypothetical situation, and a second respondent rating on the 6-point scale is obtained in response to that hypothetical. The hypothetical may vary from one patient to another and is chosen by referring to a standardized set of hypotheticals developed for the POD. (See example, Table 1.) This procedure requires that the examiner identify within the respondent's explanation the premise upon which the respondent's belief is based. That premise then becomes the basis for the proffered hypothetical, which nullifies or challenges the respondent's original belief.

The respondent's acknowledgment rating in response to the hypothetical

question is obtained and becomes the final score for that item. Only respondents who rigidly disavow the potential value of the treatment (i.e., indicate "definitely/probably not" on the 6-point rating), in the face of the hypothetical that nullifies their original premise, receive a full nonacknowledgment score (0) on that item.

NOD and NOT subtest scores make up the sum of the three items contributing to each subtest, which ranges from 0 to 6 (lower scores indicating greater nonacknowledgment). We did not conceptualize nonacknowledgment (lack of appreciation) as a general trait or cognitive style. Distortion of reality might affect a person's beliefs in one content area but not another, as when a patient recognizes the presence of psychiatric symptoms but denies (for pathological reasons) the potential for psychiatric treatment to be of possible benefit. Therefore, no total POD score was used in the development of the instrument. About 10 to 20 min is required for administration of the POD.

Thinking Rationally About Treatment (TRAT)

The TRAT provides research data relevant to psychological inferences that are made in legal inquiries applying the rational manipulation or reasoning standard for competence to consent to treatment (Appelbaum & Grisso, 1988, 1995; Grisso & Appelbaum, 1993). It assesses the quality of a person's cognitive functions that are employed in the process of deciding among alternative treatments. In contrast to the other two measures, the TRAT focuses on the person's *processing* of information.

The decision-making functions assessed by the TRAT were selected on the basis of a review of functions that are included with some consistency in various models of decision making or problem solving (e.g., Goldfried & D'Zurilla, 1969; Hogarth, 1987; Janis & Mann, 1977; Spivack, Platt, & Shure, 1976; Spivack & Shure, 1974). Eight functions were selected, and methods to assess them were incorporated into two standardized procedures: the TRAT Vignette and the TRAT Tasks.

The *TRAT Vignette* includes a presentation of a brief story describing a hypothetical patient's mental or medical illness. There are three forms of the vignette (schizophrenia, depression, and ischemic heart disease), corresponding to the three patient groups in the research studies for which the TRAT was developed. The vignette ends with a description of three treatment alternatives, as well as their benefits and risks, presented orally and on printed cards that remain in front of the respondent throughout the following process of inquiry.

At the end of the vignette, respondents are asked to pretend they are assisting the hypothetical patient by recommending one of three treatment alternatives. A series of standardized questions elicits the respondent's explanation for his or her choice, providing data for scores on five of the eight cognitive functional abilities measured in the TRAT. These five functions are defined conceptually and operationally as follows:

1. Seeking Information. This function refers to a person's tendency to seek information beyond that which is provided in the disclosure of a decision-making problem. The function is assessed by a single question that offers the respondent a chance to seek further information before making a treatment selection.

2. *Consequential Thinking*. This refers to a person's consideration of the consequences of a treatment alternative when deciding whether to reject or accept that alternative (or others). Credit for this function is received if the respondent's explanation for a treatment choice manifests the use of consequences in the reasoning for the choice.

3. *Comparative Thinking*. This refers to a person's "simultaneous" processing of information about two treatment alternatives, such that they receive consideration in relation to each other, not merely as separate facts. Credit is received if the respondent's explanation for a treatment choice refers to the consequences of two alternatives in reasonably close juxtaposition.

4. *Complex Thinking*. This refers to a person's attention to the range of treatment alternatives available within a decision problem, even if only to reject them, rather than avoiding or neglecting consideration of some alternatives. Credit for this function is given if the respondent's explanation for a treatment choice manifests reference to the full range of treatment alternatives (three alternatives) offered in the vignette.

5. *Generating Consequences*. This refers to a person's ability to generate potential real-life consequences of the liabilities described in an informed consent disclosure of a treatment alternative. This is assessed with two standardized questions that ask respondents to describe ways that medical consequences (e.g., medication side-effects) presented in the vignette might influence their own everyday activities.

In the vignette inquiry, respondents are asked to choose a preferred treatment, then are asked standardized questions to elicit their explanation for their choice. This is repeated for their least preferred choice. Detailed procedures are provided for recording their responses, and the instrument offers specific criteria and decision rules for identifying evidence of consequential, comparative, and complex thinking in respondents' explanations, leading to scores of 2, 1, or 0 for each subtest.

Three additional *TRAT Tasks* complete the TRAT procedure. They are not related to the vignette, and there are no separate forms of the tasks for various patient groups.

6. *Weighting Consequences*. This refers to a person's tendency for consistent application of his or her own preferences when evaluating the desirability of the consequences of various alternatives. This is assessed with a task presented in two parts at different times in the TRAT procedure. In Part I, respondents are presented with a series of cards, each one displaying two activities of everyday life (e.g., "go to the movies" and "read a book"). All paired combinations of five activities are presented, while in each case respondents choose one of the two activities that they prefer. Later, in Part II, the five activities are displayed simultaneously, and respondents are asked to select them in the order of their preference. The two procedures create two hierarchies of preferences for the respondent. Scoring criteria give greater credit for greater similarity between the two hierarchies (greater consistency in the respondent's application of preferences across the two tasks). Scores range from 0 to 3.

7. *Transitive Thinking*. This refers to a person's ability to make logical infer-

ences about the relative quantitative relationships between several alternatives based on paired comparisons. It is assessed with three items presenting transitive problems (for example: "A is larger than B, B is larger than C; choose the largest."). Scores range from 0 to 3.

8. Probabilistic Thinking. This refers to a person's ability to distinguish correctly the relative values of numerical probabilities stated as frequencies of occurrence. It is assessed with six questions (e.g., "Janet had a bad ulcer. Of every 100 people who have operations on their stomachs to treat their ulcers, 85 are cured, 10 remain the same, and 5 get worse. Is the operation more likely to help or not help Janet?"). Scores range from 0 to 3.

The total TRAT score is the sum of the scores on the eight subtests (range, 0 to 19). About 25–30 min are required for administration of the TRAT.

Expressing a Choice (EC)

Respondents' abilities to state a treatment preference were assessed by a single item included in the TRAT. Immediately following presentation of the TRAT vignette (but before the TRAT vignette questions are given), respondents are asked which one of the three treatment options they would recommend to the hypothetical patient. A statement of any single choice is adequate to obtain full 2-point credit; ambivalence among two or all three options results in lower credit. EC data are not considered in the remainder of this report, because only a very small percentage of respondents failed to obtain full credit on the EC item (Grisso & Appelbaum, 1995).

Procedure

Pilot Studies

The final versions of the MacArthur Treatment Competence Research Instruments were developed in an iterative process involving administrations of prototypes to patients with mental illness, in the course of multiple revisions. A more controlled pilot study then was performed with small samples of patients such as those who were to be participants in the present study. The UTD results of that pilot study have been reported (Grisso & Appelbaum, 1991). (The samples in these earlier pilot studies are not part of the samples in the full study reported here.)

Quality Control for Administration and Scoring

Data were collected by 10 research assistants at three sites, with assistants at each site being supervised by a site director. A principal investigator and the project director at the Worcester site conducted extensive training sessions at each site prior to the start of data collection. This included didactic training, practice administration sessions with feedback, group scoring, and multiple checks of individual scoring. Over a period of several weeks, assistants at all sites continued to perform practice administrations and scoring, to submit these to the

main site for review and feedback, and to score protocols that had been administered and scored by other assistants (on which preliminary interrater reliability analyses were performed).

Data Collection

Details of the data collection process with participants are provided in a companion article (Grisso & Appelbaum, 1995). Generally, patients meeting study criteria for age (18–65), language (English), diagnosis, and clearance by their treating clinicians were approached within 2 to 10 days after their hospital admissions. They were provided disclosures concerning the research purposes and task, were offered $10 for participation, and were fully informed concerning the voluntary nature of their participation and the lack of relation of the project to their current treatment. Administration of the measures occurred according to a standardized sequence (clinical measures, UTD, POD, TRAT) in surroundings that minimized distractions.

Subsets of participants in the schizophrenia group (n = 27), the depression group (n = 23), and one of the community control groups (n = 28) were readministered the three experimental instruments about two weeks (range = 14–20 days) after the first administration. This second session also included readministration of the Beck Depression Inventory to the depression group and the Brief Psychiatric Rating Scale to the schizophrenia group. The second administration for the groups with mental illness was performed either in hospital or in the community, depending on whether patients had been discharged between the first and second administrations.

Interscorer Reliability

Each research site produced a pool of 20 protocols for reliability checks (UTD and TRAT with mentally ill participants), comprising 5 to 7 administrations per research assistant. All of a site's 20 protocols were scored independently by all research assistants at that site and by a "master scorer" at the central project site (Worcester). For each instrument and its items and/or subtests, correlations (intraclass for scales and kappa for items employing discrete scores) then were performed between the master scorer and each research assistant, and between research assistants within sites. All comparisons used n = 20.

Several scales did not require examination of interscorer reliability: the UTD's Element–Recognition scores and transitive thinking and probabilistic thinking on the TRAT, because their scoring is wholly objective, and POD subtest scores, because these are based on participants' own ratings.

Internal Consistency

For subtests of the UTD and POD, and for total TRAT scores, internal consistency was examined with alpha coefficients, item-to-scale correlations, and corrected alpha coefficients (recalculated with each item deleted). These were performed separately for the schizophrenia (n = 75), depression (n = 92) and

ischemic heart disease (n = 82) samples, as well as for all three hospitalized samples combined (n = 249), all three community samples combined (n = 249), and all participants combined (n = 498). (The POD was not included in analyses involving the last two of these data sets, because it was not appropriate for administration to the community samples.)

Finally, the dimensionality and internal consistency of the TRAT was examined with factor analyses for each hospitalized sample and for the community samples combined. To determine the number of components to retain, we used the minimum average partial (MAP) method (Velicer, 1976), an approach that has been shown to be more accurate in identifying the number of principal factors to be rotated than the eigenvalue or Bartlett test rules and somewhat better than the scree method (Zwick & Velicer, 1982, 1986). A statistical package developed by Gorsuch (1990) calculates MAP, estimates communalities, extracts principal axes, rotates by Varimax, and proceeds to Promax rotation when multiple factors are indicated. Normalized scores were used for all factor analyses. (Results using nonnormalized scores proved to be similar.)

Intertest Correlations

The relations between the three instruments were examined with Pearson r and point biserial correlations, as well as a principal axes factor analysis (normalized scores) entering 13 subtest scores (3 UTD subtests, 2 POD subtests, and 8 TRAT subtests) and using the MAP method and statistical package described earlier (Gorsuch, 1990). Analyses were performed using the schizophrenia and depression samples combined (n = 167). An additional analysis was performed for the schizophrenia sample alone, although it should be noted that the ratio of variables to subjects (1:5.7) for this analysis was less than is often recommended.

Test–Retest Correlations

Potential changes across time in scores on the UTD, POD, and TRAT were examined in three ways, using 2-week retest scores with participants in the subsamples noted earlier. First, scores on first and second administration were compared (by participant group) with sign tests, t tests, and Pearson r correlations. Second, in order to examine the relation of changes in scores on the experimental measures to changes in severity of disorder, we calculated the correlations (separately for each participant group) between the difference scores on each experimental measure and the difference scores on the clinical measures (BPRS and BDI). The difference scores were "partialed" (regression-adjusted, residualized) change scores, in which the first score is regressed from the second score (Cohen & Cohen, 1983).

RESULTS

Scores on all 13 subtests were distributed across the full range of possible scores. In general, however, scores on the UTD subtests were not normally

distributed; medians were nearer to the high end (positive performance) of the possible range. The exception was for the schizophrenia group, whose scores tended to be normally distributed, with the median and mean near the middle of the subtests' range of scores. Most standard deviations for UTD subtests ranged from 1.5 to 2.5 (range of scores = 0–10).

POD subtest scores also were not normally distributed, the majority scoring at the high (positive performance) end on both subtests. Total TRAT scores were more normally distributed for all participant groups, with standard deviations ranging from 2.5 to 4.1 (range of scores = 0–19).

Interscorer Reliability

UTD

Comparing each of the 10 scorers to the "master scorer" for the UTD subtests (Uninterrupted–Paraphrase and Element–Paraphrase), intraclass correlations for subtest scores all were above .84 (M = .91). Kappa correlations for individual UTD items were .60 or above for 90% of the comparisons, and .70 or above for 74% of the comparisons. Among scorers within data collection sites, intraclass correlations for UTD subtests all were above .88 (M = .91); kappa correlations for individual UTD items were .60 or above for 88% of the comparisons, and .70 or above for 76% of the comparisons. All 156 interscorer correlations referred to above were statistically significant ($p < .01$).

TRAT

Comparing each of the 10 scorers to the master scorer for the TRAT total score, intraclass correlations all were above .85 (M = .91). Kappa correlations for individual TRAT subtests were .60 or above for 76% of the comparisons. One scorer's master-scorer comparisons on two subtests, and another's on one subtest, fell below the accepted level of statistical significance. (These constituted 7% of the master-scorer kappa comparisons for the TRAT.) Similarly, among scorers within data collection sites, intraclass correlations for total TRAT scores all were above .88 (M = .90), kappa correlations for individual TRAT subtests were .60 or above for 77% of the comparisons, and one of the 18 comparisons failed to achieve the accepted level of statistical significance.

Although these levels of correlation overall were considered quite good, it was decided for the sake of caution to have the project's master scorer score all TRAT protocols for the project and to use those scores as the TRAT data base for all further analyses in the study.

Internal Consistency

UTD

Table 2 shows alpha coefficients and corrected item–scale correlations (when an item is removed) for the Uninterrupted–Paraphrase and Element–Paraphrase subtests, separately for analyses by various experimental groups. Item–scale

Table 2. Alpha and Item–Scale Coefficients for Subtests of Understanding Treatment Disclosure (UTD)

Subtests and groups	Alpha	Range (and average) of item–scale correlations	Outliers
Uninterrupted–Paraphrase			
Schizophrenia	.83	.53–.78 (.64)	
Depression	.75	.54–.57 (.52)	Item 4 (.36)
Angina	.62	.39–.53 (.38)	Item 1 (.23)
			Item 5 (.26)
All hospitalized	.75	.43–.59 (.52)	
All community	.70	.36–.53 (.46)	
Element–Paraphrase			
Schizophrenia	.85	.63–.68 (.66)	
Depression	.67	.32–.51 (.44)	
Angina	.66	.36–.53 (.42)	
All hospitalized	.77	.45–.62 (.54)	
All community	.55	.23–.38 (.31)	

correlations for what appear to be "outlier" items are listed separately, as well as corrected alpha coefficients. Alpha coefficients and item–scale correlations were best for the schizophrenia group, but relatively good for all other groups except the community samples on Element–Paraphrase. Finally, the three subtests were substantially correlated with each other, somewhat more so for the hospitalized participants than for the community samples (see Table 5).

POD

Using item and subtest scores of all three hospitalized groups (N = 249), alpha coefficients for Nonacknowledgment of Disorder (NOD) and Nonacknowledgment of Treatment Potential (NOT), respectively, were .80 and .67. Corrected item–total correlations were good for NOD (.70, .59 and .66), whereas NOT items were correlated only modestly with their total subtest scores (.43, .60 and .45). Scores on the two subtests, however, were poorly correlated with each other (.23).

TRAT

Several of the TRAT subtests are based on a single scored response. Therefore, each subtest was treated as an item contributing to the overall TRAT as a scale. Table 3 shows alpha coefficients and item–scale correlations for the TRAT, separately for analyses by various experimental groups. Item–scale correlations for what appear to be outlier items are listed separately, as well as corrected alpha coefficients. As can be seen in Table 3, these statistics indicated some internal inconsistency among various subtests (items) of the TRAT; it was questionable whether the subtests were tapping a general ability construct in common.

To examine the TRAT's complexity, we performed factor analyses using the

Table 3. Alpha and Item–Scale Coefficients for Subtests of Rational Thinking About Treatment (TRAT)

Group	Alpha	Range and average of item–scale correlations	Outliers[a]	Alpha with outlier removed
Schizophrenia	.74	.38–.67 (.42)	WEIGH (.27)	.74
Depression	.58	.32–.55 (.33)	WEIGH (.08)	.59
			SEEK (.14)	.58
Angina	.39	.20–.51 (.25)	GENER (.01)	.44
			TRANS (.03)	.43
All hospitalized	.68	.26–.49 (.37)		
All community	.55	.26–.39 (.34)	WEIGH (.00)	.58
			SEEK (.11)	.53
			GENER (.19)	.51

[a] WEIGH = Weighting Consequences; SEEK = Seeking Information; GENER = Generating Consequences; TRANS = Transitive Thinking; PROBA = Probabalistic Thinking.

minimum average partial (MAP) method and the statistical package by Gorsuch (1990) (see Method section). Analyses were performed separately for each of the hospitalized samples, and for the combined community samples, entering all eight subtests. The MAP method selected only one component to be retained on each of the analyses. Upon inspection of the resulting analyses, however, two subtests—Seeking Information and Weighting Consequences—manifested relatively low factor–variable correlations fairly consistently across samples (generally ranging from .01 to .30).

Therefore, we performed additional factor analyses (using the same factor selection and rotation methods described earlier) entering only six of the TRAT variables (deleting Seeking Information and Weighting Consequences). For each hospitalized sample and for the community samples combined, a single factor solution again was indicated. The results, shown in Table 4, indicate adequate factor–variable correlations for the single factor with the schizophrenia and depression samples, but less than adequate correlations for some subtests in the

Table 4. Correlations of Six Subtests of TRAT ("TRAT-2") with Single Factor from Factor Analysis, for Hospitalized Samples and the Combined Community Samples

	Samples			
Subtests	Schizophrenia (n = 75)	Depression (n = 92)	Angina (n = 82)	Community (n = 249)
Consequential Thinking	.84	.61	.67	.67
Comparative Thinking	.72	.62	.75	.71
Complex Thinking	.51	.34	.40	.52
Transitive Thinking	.37	.43	−.03	.25
Probabilistic Thinking	.40	.45	−.05	.30
Generating Consequences	.37	.39	.11	.19
Percent of variance	32	24	20	23

analysis for the angina group. In later analyses, TRAT scores based on six subtests alone (deleting Seeking Information and Weighting Consequences) are referred to as *TRAT-2* scores.

Relations Between Instruments

Table 5 shows correlations between UTD subtests, POD subtests, and overall TRAT and TRAT-2 scores. These are shown separately for hospitalized and for community participants. UTD subtest scores were correlated modestly with TRAT and TRAT-2 scores for the hospitalized samples, but less so among the community samples. Scores on the two POD subtests correlated poorly with both UTD and TRAT scores.

Factor analyses entering all 13 subtests were performed for the schizophrenia group alone and the schizophrenia and depression groups combined (using normalized scores and the MAP and rotation methods described earlier). A two-factor solution was indicated in both cases, and Table 6 shows the factor–variable correlations for the Promax factor structure. In general, the first factor is represented best by the UTD subtests and the second by TRAT subtests derived from the TRAT vignette procedure (Consequential, Comparative, and Complex Thinking). The TRAT's Generating Consequences and both Transitive and Probabilistic Thinking were modestly related to both factors, whereas the two POD subtests were related to neither factor.

Test–Retest Correlations

Table 7 shows mean scores for the three retested subsamples at first and second testing on the UTD subtests, the POD subtests, and the TRAT-2. Changes

Table 5. Correlations Between Experimental Measures[a]

	UTD			POD		TRAT	
Measures[b]	UD-PR	ED-PR	ED-RC	NOD	NOT	TRAT	TRAT-2
UTD							
UD-PR		.62	.40	NA[c]	NA	.18	.23
ED-PR	.72		.47	NA	NA	.26	.29
ED-RC	.55	.61		NA	NA	.30	.30
POD							
NOD	.22	.15	.23			NA	NA
NOT	.09	.06	−.08	.23			
TRAT							
TRAT	.34	.46	.40	.14	−.04		
TRAT-2	.32	.50	.42	.12	.07		

[a] Correlations to the right of the diagonal are for community participants (n = 249) and to the left are for hospitalized participants (n = 249).
[b] UD-PR = Uninterrupted–Paraphrase; ED-PR = Element–Paraphrase; ED-RC = Element–Recognition; NOD = Nonacknowledgment of Disorder; NOT = Nonacknowledgment of Treatment Potential.
[c] The POD was not appropriate for administration to community participants.

Table 6. Correlations of 13 Subtests of UTD, POD, and TRAT with Two Factors, for Two Mental Illness Samples Combined and for Schizophrenia Sample

	Sample			
	Schizophrenia and depression (n = 167)		Schizophrenia (n = 75)	
Subtests	Factor I	Factor II	Factor I	Factor II
UTD: Uninterrupted–Paraphrase	.79	.49	.86	.12
UTD: Element–Paraphrase	.77	.52	.86	.25
UTD: Element–Recognition	.61	.33	.77	.08
TRAT: Transitive Thinking	.48	.45	.53	.28
TRAT: Probabilistic Thinking	.49	.43	.50	.26
TRAT: Consequential Thinking	.44	.70	.30	.84
TRAT: Comparative Thinking	.42	.67	.27	.74
TRAT: Complex Thinking	.24	.38	.07	.52
TRAT: Seeking Information	.19	.37	.15	.59
TRAT: Generating Consequences	.40	.40	.34	.36
TRAT: Weighting Consequences	.23	.23	.34	.22
POD: Disorder	−.10	.00	−.11	.12
POD: Treatment	−.09	.03	−.16	.22

in severity of symptoms also are shown on the BPRS (for the schizophrenia group) and the BDI (for the depression group).

For the UTD subtests and the TRAT-2 alike, correlations between first and second administration were highly significant and generally in the range of .50–.80, with the exception of the Uninterrupted–Paraphrase. A significant decrease in BDI (depression) scores was observed across time for the depression group; there was commensurate improvement in performance on two of the three UTD scales, while TRAT-2 means were not significantly different at the two administrations. In contrast, neither the schizophrenia group's BPRS severity scores nor performance on the UTD or TRAT-2 changed significantly between the two administrations.

This result for the schizophrenia group was difficult to interpret, in that the mean BPRS, UTD, and TRAT scores do not reflect the fact that some patients might have experienced increases in symptom severity, while others experienced decreases, thus masking changes in performance related to changes in symptom severity. Table 8, therefore, shows correlations between difference scores on the experimental subtests and difference scores on the clinical measures. In general, among schizophrenia patients, the amount of change in performance on the UTD and TRAT-2—but not on the POD subtests—was correlated with the amount of change in clinical severity of symptoms (on the BPRS or BDI). Among the depressed patients, however, changes in POD scores alone were related to changes in symptom severity.

Table 7. Mean Scores at First and Second Testing (Approximately Two-Week Interval) on UTD, POD, TRAT-2, and Measures of Symptom Severity (BPRS, BDI) for Retested Subsamples

Samples and measures	Testing session		Pearson r[a]
	First	Second	
Schizophrenia (n = 27)			
UTD: Uninterrupted–Paraphrase	4.3	5.3	(.30)
UTD: Element–Paraphrase	5.8	6.2	.55
UTD: Element–Recognition	6.4	6.9	.60
POD: Nonacknowledgment of Disorder	4.6	5.3[b]	.56
POD: Nonacknowledgment of Treatment Potential	4.4	4.8	.90
TRAT-2	7.3	7.8	.66
BPRS	45.7	43.6	.57
Depression (n = 23)			
UTD: Uninterrupted–Paraphrase	6.9	7.8[b]	.45
UTD: Element–Paraphrase	8.3	8.5	.52
UTD: Element–Recognition	8.5	9.0[b]	.71
POD: Nonacknowledgment of Disorder	5.8	5.9	.59
POD: Nonacknowledgment of Treatment Potential	5.3	5.3	.48
TRAT-2	9.7	9.9	.68
BDI	31.8	17.5[b]	.44
Community (n = 28)			
UTD: Uninterrupted–Paraphrase	6.8	8.0[b]	.47
UTD: Element–Paraphrase	8.9	9.0	.56
UTD: Element–Recognition	9.2	9.2	.80
TRAT-2	10.6	10.7	.68

[a] All correlations are significant, $p < .001$, unless shown in parentheses.
[b] Means are significantly different, $p < .01$ (t-tests).

DISCUSSION

The research instruments described in this report were developed conceptually to assess the functional abilities of patients to participate meaningfully in making decisions about treatment for mental or general medical disorders. The present report does not address directly whether the instruments measure the abilities that they purport to measure. Evidence of construct validity (e.g., relations of scores to psychopathology and a measure of general intellectual functioning) is offered in a companion article (Grisso & Appelbaum, 1995). Here we examined only the psychometric properties of the instruments, especially their reliability.

Evidence suggests that most subtests in the three instruments can be scored reliably by nonprofessionals (postbaccalaureate research assistants) if they are given adequate training. Reliability in the scoring of certain TRAT subtests, however, may require special care and consideration in future studies using this instrument. In addition, it should be noted that the present study did not address the reliability with which the instruments could be administered (i.e., whether different administrators of the procedures would obtain similar performances from patients).

Table 8. Correlations Between Difference Scores (First and Second Testing) on Experimental Measures and on Measures of Severity of Symptoms, for Schizophrenia and Depression Samples

Groups and subtests	Symptom severity measures	
	BPRS	BDI
Schizophrenia (n = 27)		
UTD: Uninterrupted–Paraphrase	−.22	−.39*
UTD: Element–Paraphrase	−.45**	−.58**
UTD: Element–Recognition	−.54**	−.49**
POD: Nonacknowledgment of Disorder	−.30*	−.02
POD: Nonacknowledgment of Treatment Potential	−.20	.11
TRAT-2	−.60***	−.46*
Depression (n = 23)		
UTD: Uninterrupted–Paraphrase	−.07	.11
UTD: Element–Paraphrase	−.01	.11
UTD: Element–Recognition	−.13	.06
POD: Nonacknowledgment of Disorder	−.33*	−.23
POD: Nonacknowledgment of Treatment Potential	−.18	−.48**
TRAT-2	.12	.18

Note. Negative valences indicate that scores on the two measures being compared vary inversely. Higher scores on the experimental measures indicate "better" performance; therefore, their inverse correlations with symptom severity indicate that performance improves as symptom severity is reduced.

* $p < .05$.
** $p < .01$.
*** $p < .002$.

Concerning internal consistency, substantial relations were found among items within UTD subtests for most groups. Poorer alpha coefficients for Element–Paraphrase data from community samples may have been due to the fact that many of their scores were near the "ceiling" for the instrument, which could have attenuated the coefficients that possibly could be attained. Substantial relations also were found among UTD subtests themselves. This suggests internal consistency of the UTD (and of its subtests) as a measure of performance on a task designed to assess patients' understanding of information relevant for their participation in making treatment decisions. It seems likely that, in future studies, either the Uninterrupted–Paraphrase or the Element–Paraphrase subtests could be used alone to represent the understanding construct that guided the development of the UTD.

For the POD instrument, the Nonacknowledgment of Disorder subtest showed adequate internal consistency, but Nonacknowledgment of Treatment Potential manifested marginal item–subtest correlations. Moreover, the low correlation between the two subtests suggests that the POD does not measure a single trait or cognitive tendency related to nonacknowledgment of the relevance of diagnostic or treatment-related information for oneself. In other words, whether participants acknowledged or failed to acknowledge aspects of their disorder was not predictive of their acknowledgment or failure to acknowledge the likelihood that they might benefit from treatment.

In this sense, the POD and its two subtests do not seem to have the properties of "scales." The POD is best seen as a set of interview screening questions, with nonacknowledgment on any one of them raising a concern about a respondent's unrealistic rejection of the relevance of diagnostic or treatment information for his or her own circumstances. The subtest scores may be interpreted as quantitative indicators that one or more items have "detected" nonacknowledgment. The fact that the two parts of the POD correlated poorly with each other, however, suggests that the scores should not be interpreted by themselves as signifying any particular trait (e.g., generalized "denial").

The TRAT as originally conceptualized appears to have two "outlying" subtests—Weighting Consequences and Seeking Information—that do not relate to the central component with which the other six TRAT subtests are associated. For this reason, we recommend that future research might profitably use only the other six subtests, which we have labeled *TRAT-2*. The Weighting Consequences and Seeking Information subtests may still represent abilities that could influence reasoning, but they may tap cognitive components different from those represented in the primary factor with which the other TRAT subtests are associated. Results offered in a companion article (Grisso & Appelbaum, 1995), however, suggest that their inclusion or exclusion in computing a total TRAT score will not produce appreciably different results for most research purposes.

The single-factor solution for the TRAT-2 was not as satisfactory for the angina group and for the combined community groups as for the schizophrenia and depression groups (see Table 4). This may have been due to the better performance overall of the former groups, so that variation in their scores was more likely to reflect ambient error in the instrument than systematic individual differences in ability.

Concerning the relations between measures, evidence from the intercorrelations (Table 5) and factor analyses (Table 6) indicate that the UTD and the TRAT tap different but somewhat related abilities. Consistent with this interpretation, even items loading highest on one factor tended to load at least moderately on the other factor. For example, for the combined schizophrenia and depression groups, the UTD subtests defined the first factor but also were related substantially to the second factor, which was defined especially by the TRAT vignette subtests. In contrast, the two POD subtests (measuring acknowledgment of one's disorder and the potential value of treatment) were unrelated to both factors. Together with the results in Table 5, this suggests that the POD subtests tap constructs not associated with the other two measures.

If we presume that the three measures assess functions associated with three of the four main legal standards for competence, the results suggest a certain validity to the law's notion of multiple standards for competence. Strengths or deficits in one type of ability do not necessarily predict one's status on other abilities that are related conceptually to participation in decisions about one's treatment.

Test–retest scores for the community subsample suggest a moderate degree of consistency in scores across a 2-week interval. Somewhat less consistency was observed for the mentally ill subsamples, probably because there were changes in

psychiatric status for some patients and not for others. This would be expected to produce lower test–retest correlations for a given patient group, presuming some relationship between clinical status and manifestation of abilities assessed by the measures. The experimental measures' sensitivity to such changes in clinical status is suggested by the analysis of correlations between difference scores on these two types of measures.

The relation between changes in clinical status and changes in performance on the UTD and TRAT was stronger for the schizophrenia group than for the depression group. One possible explanation is that the majority of depression respondents performed within a relatively truncated (high) range of scores on these two measures, whereas scores of the schizophrenia group were more widely distributed. With many of the depression respondents' scores near the ceiling on the instruments, conditions may have attenuated the correlation that could be attained.

In general, we believe that the test–retest data support the notion that the UTD Element–Paraphrase and Element–Recognition subtests, as well as the TRAT, are influenced by relatively stable cognitive abilities, but that they are also sensitive to changes in mental status that would influence the manner in which these cognitive abilities are manifested in performance at a given time. This interpretation is consistent with the current conceptualization of legal competence to make treatment decisions. The law does not conceptualize competence or incompetence as a static condition, but rather as reflecting the person's present functional status, which can change in relation to changing clinical or environmental conditions.

We do not recommend the experimental measures described here for routine clinical use in evaluating patients for decisional abilities related to competence to consent to treatment. They require considerable time and effort for administration and scoring, more than is feasible for most clinical situations. The measures, however, appear to have adequate reliability and stability for use in research that requires indices of abilities related conceptually to patients' legal competence to make treatment decisions.

REFERENCES

Appelbaum, P., & Grisso, T. (1988). Assessing patients' capacities to consent to treatment. *New England Journal of Medicine, 319,* 1635–1638.

Appelbaum, P., & Grisso, T. (1992). *Manual: Perceptions of disorder.* Worcester MA: University of Massachusetts Medical Center.

Appelbaum, P., & Grisso, T. (1995). The MacArthur Treatment Competence Study. I: Mental illness and competence to consent to treatment. *Law and Human Behavior, 19,* 105–126.

Appelbaum, P., Lidz, C., & Meisel, A. (1987). *Informed consent: Legal theory and clinical practice.* New York: Oxford University Press.

Beck, A. (1978). *Beck Depression Inventory.* San Antonio TX: Psychological Corporation.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Cutter, M. A. G., & Shelp, E. E. (1991). *Competency: A study of informal competency determinations in primary care.* Dordrecht: Kluwer Academic Publishers.

Goldfried, M., & D'Zurilla, T. (1969). A behavioral-analytic model for assessing competence. In C. D. Spielberger (Ed.), *Current topics in clinical and community psychology* (Vol. 1, pp. 151–196). New York: Academic Press.

Gorsuch, R. (1990). *UniMult guide*. Altadena, CA: Fuller Theological Seminary.

Grisso, T. (1986). *Evaluating competencies: Forensic assessments and instruments*. New York: Plenum.

Grisso, T., & Appelbaum, P. (1991). Mentally ill and non-mentally ill patients' abilities to understand informed consent disclosures for medication. *Law and Human Behavior, 15*, 377–388.

Grisso, T., & Appelbaum, P. (1992). *Manual: Understanding treatment disclosures*. Worcester MA: University of Massachusetts Medical Center.

Grisso, T., & Appelbaum, P. (1993). *Manual: Thinking rationally about treatment*. Worcester, MA: University of Massachusetts Medical Center.

Grisso, T., & Appelbaum, P. (1995). The MacArthur Treatment Competence Study. III: Abilities of patients to consent to psychiatric and medical treatments. *Law and Human Behavior, 19*, 149–174.

Grunder, T. (1978). Two formulas for determining the readability of subject consent forms. *American Psychologist, 33*, 773–775.

Hogarth, R. (1987). *Judgement and choice: The psychology of decision*. New York: John Wiley.

Janis, I., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and communication*. New York: Free Press.

Overall, J. (1988). The Brief Psychiatric Rating Scale (BPRS): Recent developments in ascertainment and scaling. *Psychopharmacology Bulletin, 24*, 97–99.

President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research (1982). *Making health care decisions*. Vol. 1. Washington, DC: U.S. Government Printing Office.

Rivers v. Katz, 112 A.D.2d 926, 491 N.Y.S.2d 1011 (1985).

Rogers v. Commissioner, Dep't of Mental Health, 390 Mass. 489, N.E.2d 308 (1983).

Roth, L., Meisel, A., & Lidz, C. (1977). Tests of competency to consent to treatment. *American Journal of Psychiatry, 134*, 279–284.

Spivack, G., Platt, J., & Shure, M. (1976). *The problem solving approach to adjustment*. San Francisco, CA: Jossey-Bass.

Spivack, G., & Shure, M. (1974). *Social adjustment of young children: A cognitive approach to solving real-life problems*. San Francisco, CA: Jossey-Bass.

Tepper, A., & Elwork, A. (1984). Competence to consent to treatment as a psycholegal construct. *Law and Human Behavior, 8*, 205–223.

Velicer, W. (1976). The relation between factor score estimates, image scores and principal component scores. *Educational and Psychological Measurement, 36*, 149–159.

Zwick, W., & Velicer, W. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research, 17*, 253–269.

Zwick, W., & Velicer, W. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432–442.