



NIMH-MATRICES survey on assessment of neurocognition in schizophrenia

Robert S. Kern^{a,b,*}, Michael F. Green^{a,b}, Keith H. Nuechterlein^{a,c}, Bi-Hong Deng^a

^a*UCLA Department of Psychiatry and Biobehavioral Sciences, Los Angeles, California, United States*

^b*Department of Veterans Affairs VISN 22 Mental Illness Research, Education, and Clinical Center (MIRECC), United States*

^c*UCLA Department of Psychology, Los Angeles, California, United States*

Received 3 September 2004; accepted 3 September 2004

Available online 12 October 2004

Abstract

The NIMH-Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) initiative requires, among other things, the establishment of a reliable, valid, and consensus-derived method of assessing cognition. The derived battery will provide a standardized way to assess the effects of cognition-enhancing agents across clinical trials. To this end, the first of six consensus-oriented conferences was held April 2003. The goals were twofold: (a) To select which cognitive constructs to measure in a consensus battery, and (b) to select which criteria to use in evaluating tests for inclusion in the battery. Based on consultation with experts on the RAND Panel Method, 74 experts were invited to participate in a pre-meeting survey to provide information relevant to decisions on the cognitive battery. The survey included sections on reliability, validity, test administration, norms and interpretation of tests, cognitive domains and their integration, battery duration, and overall importance of test qualities. For selection of cognitive targets, the results showed that experts ranked executive functions, attention/vigilance, memory processes, and problem-solving ability highest. For test qualities, the experts ranked test–retest reliability, good coverage of key individual cognitive constructs, and comparable alternate forms highest. This article presents the results of the pre-conference survey that was the first step in the RAND process towards development of the NIMH-MATRICES consensus battery to assess cognition in schizophrenia.

© 2004 Elsevier B.V. All rights reserved.

Keywords: NIMH-MATRICES; Neurocognition; Schizophrenia

1. Introduction

One clearly identified obstacle for the development of cognition-enhancing agents in schizophrenia is the absence of a reliable, valid, and consensus-derived cognitive assessment method. With this obstacle in mind, the goals of the first MATRICS

* Corresponding author. VA Greater Los Angeles Healthcare System (MIRECC 210 A), 11301 Wilshire Boulevard, Los Angeles, CA 90073, United States. Tel.: +1 310 478 3711x49229; fax: +1 310 268 4056.

E-mail address: rkem@ucla.edu (R.S. Kern).

consensus-oriented conference (held in April, 2003) were twofold: (a) To select which cognitive constructs to measure, and (b) to select which criteria to use in evaluating specific tests for inclusion in the battery. Prior to the meeting, a group of experts were invited by the MATRICS Neurocognition Committee to participate in a survey and attend the consensus conference. The inclusion of survey data in making consensus decisions was based on consultation with experts on the RAND Panel Method (Fitch et al., 2001), who emphasized the need to gather information systematically from key stakeholders in the process. The selection of experts (listed at the end of this article) was done by members of the MATRICS Neurocognition Committee. Based on results of the survey, discussions at the meeting, and a review of factor analytic studies (see Nuechterlein et al., 2004), the MATRICS Neurocognitive Committee made decisions about cognitive domains and essential test criteria. This article will present the results of the pre-conference survey.

As shown in Table 1, 74 experts were invited to participate. Sixty-two agreed to participate plus an

additional 6 members from the MATRICS Neurocognition Committee, yielding a total of 68 completed surveys. The experts were primarily from academia and included persons with varying areas of expertise. Members of other key sectors with vested interest in the process were represented as well. These included members from government (NIMH), pharmaceutical industry, and test publishers.

The survey was written by the authors of this article with input from the MATRICS Neurocognition Committee. The final survey included sections on reliability, validity, test administration, norms and interpretation of tests, cognitive domains and their integration, battery duration, and overall importance of test qualities. The interview included both open-ended questions and Likert-scale ratings. The majority of the surveys were conducted by phone interview to maximize qualitative input from experts. A few surveys were completed independently by experts and returned via e-mail. The phone interviews took roughly 45–50 min to conduct. A summary of the results from each section of the survey follows.

Table 1
Characteristics of experts surveyed

Expert characteristics	Total
<i>No. of experts</i>	
Invited	74
Agreed and completed	62
Neurocognition committee members	6
Total	68
<i>Sectors represented</i>	
Academia	53
Government	3
Industry	7
Other	5
<i>Areas of expertise</i>	
Cognitive performance deficits in schizophrenia	42
Cognitive neuroscience	34
Clinical trials methodology	27
Neuropharmacology	25
Outcome assessment	24
Cognitive science	22
Psychometrics/test development	21
Biostatistics	4
Other	2

* Participants could choose more than one.

2. Reliability

This section addressed two types of reliability that are especially relevant to clinical trials designs: test–retest reliability and alternate form reliability.

2.1. Test–retest reliability

The experts were asked to comment on: (a) the importance of test–retest reliability in evaluating cognitive measures for use in studies with clinical trials designs, and (b) the influence of practice effects in repeated-measure designs. As shown in Fig. 1, most experts considered test–retest reliability to be an essential criterion for test selection for this purpose. Many experts offered the well-established rationale that the upper limit of validity is dependent on a test's reliability (Anastasi, 1988). For clinical trial designs, the biggest concern was that increased error variance associated with lower test–retest reliability would make detection of group differences difficult and would necessitate increased sample sizes. Most experts agreed that tests needed to have

high test–retest reliability, but it need not be extremely high. Unusually high test–retest reliability over a substantial period of time can, in some cases, reflect a construct that is so stable that it would unlikely change in response to treatment (e.g., verbal intelligence).

In most cases, practice effects were seen as tolerable for clinical trials as long as performance was not limited by scaling issues (e.g., ceiling effects). In studies that use parallel groups, inferences can be made in the presence of practice effects as long as no one group approaches ceiling. However, in tests that have particular solutions (e.g., Wisconsin Card Sorting Test [WCST]), practice effects pose a potentially more serious problem because once subjects discover the key strategy or underlying principle, it dramatically affects future performance (Lezak, 1995).

2.2. Alternate form reliability

The experts were asked to comment on the advantages and disadvantages in the use of alternate forms for tests affected by simple repetition. They were also asked to comment on the level of comparability necessary between forms, and the number of forms desirable for a clinical trials design. As shown in Fig. 2, the experts viewed highly comparable alternate forms as essential for tests with substantial practice effects. This was seen as highly desirable for tests of memory with specific material to be learned (e.g., word lists,

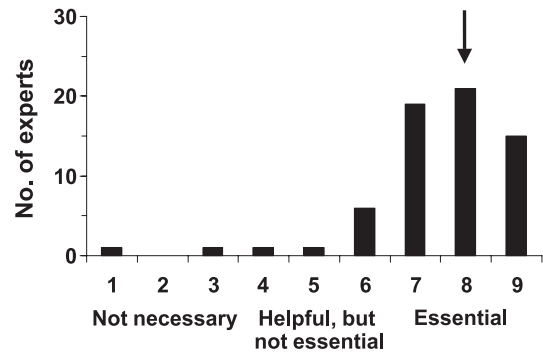


Fig. 2. Alternate form reliability. How important is it for cognitive tests that are affected by repetition to have alternate forms with high comparability?

short passages) and tests of executive functioning with a key rule or concept to discover (e.g., WCST). For speed of processing and attention/vigilance tests, measures typically associated with minor practice effects (Nuechterlein et al., 1998; Wechsler, 1997), alternate forms were not seen as necessary. On the other hand, some experts cautioned against the use of alternate forms in clinical trials. They argued that unless the forms are highly correlated and have similar means and standard deviations the benefits of minimizing practice effects may be offset by adding another source of error. In situations that called for using alternate forms, the experts generally preferred the number of alternate forms to equal the number of assessment points (e.g., typically three to five in clinical trials designs).

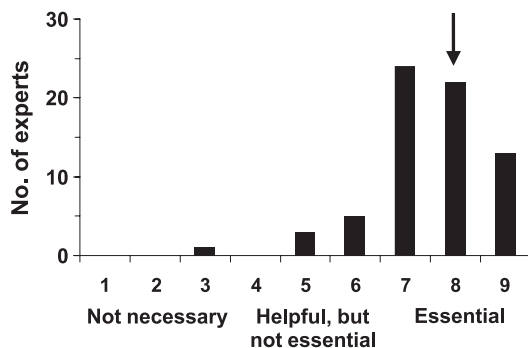


Fig. 1. Test–retest reliability. How important is it in a clinical trials situation for the cognitive test to have high short-term test–retest reliability?

3. Validity

This section addressed three types of validity: (a) predictive validity for functional outcome, (b) predictive validity for symptoms, and (c) “neuropharmacological” validity (i.e., links between cognitive test performance and underlying neural systems). In addition, the experts were asked about the importance of diagnostic specificity in selecting tests for the battery.

For predictive validity for functional outcome, experts were asked to provide a rating for the overall importance of this form of validity, and were asked to consider: (a) how much weight

should be given to the strength of the relationships between a particular cognitive measure and functional outcome, (b) to what extent did data for a test need to be longitudinal rather than solely cross-sectional, and (c) whether data are needed from a specific test within a cognitive domain, or whether information on the domain alone was sufficient. As shown in Fig. 3, predictive validity for functional outcome received a median rating of 6.0, falling in the upper end of the “helpful, but not essential” classification category. Experts identified several problems that limit the value of predictive validity for functional outcome. These included: (a) problems associated with shared “method variance” between neurocognitive and functional outcome measures such as social problem-solving tests (Bellack et al., 1999), (b) the cross-sectional nature of much of the existing literature and therefore lacking in prospective longitudinal ties (Green, 1996; Green et al., 2000; but see Green et al., 2004), and (c) the fact that many factors outside the realm of neurocognition affect functional outcome such as the amount and availability of rehabilitation and social support. Some experts pointed out that in the context of the search for cognition-enhancing agents, a critical question not yet addressed in the existing literature is whether change in neurocognition leads to change in functional outcome.

For predictive validity for psychiatric symptoms, the experts were asked to consider to what extent a cognitive measure included in the consensus battery should be linked to symptom dimensions

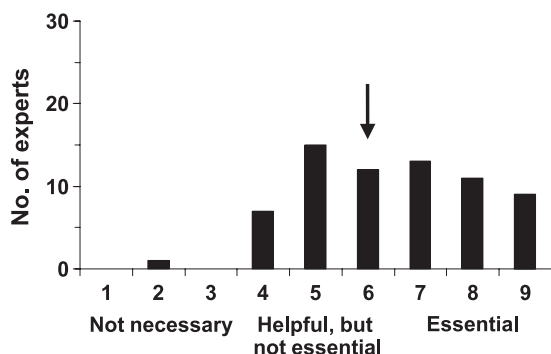


Fig. 3. Predictive validity for functional outcome. How important is it for the tests to be significantly related to the functioning level of patients?

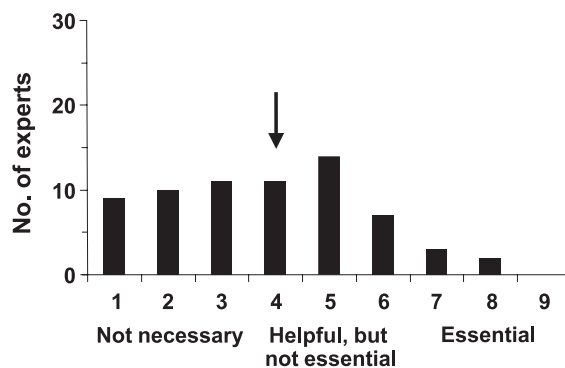


Fig. 4. Predictive validity for schizophrenic symptoms. How important is it for the tests to be related to the clinical psychopathology of patients?

in schizophrenia. As shown in Fig. 4, most experts saw this criterion as either “unnecessary” or “helpful, but not essential.” The predominant view was that cognition should be viewed as an illness dimension that is important in its own right, but which overlaps modestly with some clinical aspects of the illness, primarily negative symptoms.

In the last question on validity, experts were asked to consider how much weight should be given to the empirical database linking cognitive deficits to an underlying neural system. As shown in Fig. 5, there was considerable spread on this issue with the median rating falling in the “helpful, but not essential” classification category. The most common response was that neurocognitive tests do not have neural system specificity, and the neural systems underlying cognitive performance are not adequately understood at this time. For this reason, some argued that a linkage between a neuropharmacological system and a cognitive deficit should not be a requirement at this stage. Strongest support for neuropharmacological validity came from some experts in cognitive neuroscience who thought that there was sufficient information at this time to select tests based on their ties to neurotransmitter systems (e.g., sensory gating tied to activity in the cholinergic system).

Lastly, experts were asked about the importance of having tests that measure deficits specific to schizophrenia. Most experts did not think that this criterion was necessary (median rating=3.0; “not necessary”). It was argued that the cognitive con-

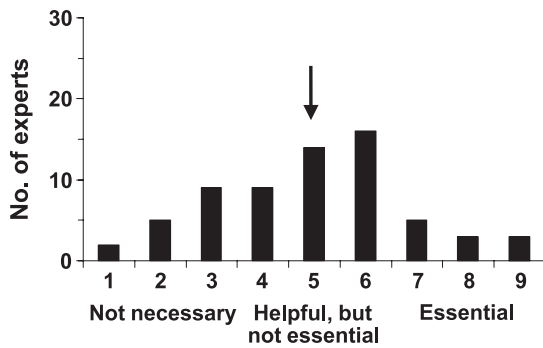


Fig. 5. “Neuropharmacological” validity. How important is it for a test to have well-established links to neuropharmacological systems involved in cognitive deficits in schizophrenia?

struct need only be relevant to schizophrenia, not diagnostically specific to it. As expressed by one expert, “The brain is not organized by diseases. Similar cognitive deficits are found among many differing neurological disorders.”

In summary, experts viewed relationships between cognitive performance and functional outcome to be the most important form of validity in construction of a cognitive battery for clinical trials. Demonstrated relationships to certain neuropharmacological systems were viewed as next most important, but believed to be largely unknown at this time. Demonstrated relationships to symptom dimensions and specificity to schizophrenia were not viewed as necessary criteria.

4. Test administration

In this section, experts were asked about training procedures for testers who will administer the neurocognitive battery, and about their preference for paper-and-pencil versus computer-based measures.

4.1. Required specific training for test administrators

Most experts indicated that it was difficult to estimate the amount of time needed for training without knowing the selection of measures included in the neurocognitive battery. The median estimate was 2 days (range=4–8 hours to 4 days). Experts who thought in terms of paper-and-pencil tests tended to estimate longer training times than those emphasizing computerized tests.

4.2. Paper-and-pencil versus computerized administration

Experts were asked about the advantages and disadvantages of paper-and-pencil versus computerized test administration. Some argued that reliability and validity should determine test selection, not mode of administration. The median rating reflected no clear preference among the experts (median=5.0). Several issues were raised. Some experts noted that the mode of administration interacts with subject preferences. Some patients are more comfortable with computers; others prefer human contact. It was also noted that decisions about mode of administration often depend on the nature of the cognitive test (e.g., visual vigilance is best measured with a computerized Continuous Performance Test). Those arguing for paper-and-pencil measures identified two primary advantages: (a) subjects are better able to attain optimal test performance via establishment of clinical rapport, and (b) interpretation of test data is enriched through access to qualitative observation of test behavior. Those arguing for computerized administration noted that regardless of whether administration is paper-and-pencil or computerized, the examiner needs to be present to monitor subjects’ understanding of, cooperation with, and engagement with the task. Computerized tests also allow standardization of administration and scoring to be more easily maintained across testers and sites. A number of experts, regardless of position, noted that computerized tests are frequently associated with loss of data due to computer malfunction or problems in saving, retrieving, and reading data files.

5. Norms and interpretation of tests

This section of the survey addressed issues related to the importance of test norms to interpret results. The first subsection addressed the overall importance of having test norms. The second subsection addressed the importance of having norms for all tests included in the battery drawn from the same reference group (i.e., co-norming). The third subsection addressed the importance of having normative data from demographic subgroups.

5.1. Test norms—overall role

Experts were asked to provide their opinion about the importance of being able to measure the level of deficit that a particular score represents compared to the general population, and whether it falls within normal limits (before and after intervention). In general, experts saw this criterion as “essential,” with a median rating across experts of 7.0. Several points were raised. First, norms have the advantage of communicating findings to a broader audience and convey clinical meaningfulness. Second, norms should be drawn from a large sample with adequate representation of age, education, gender, and ethnicity. Third, the utility of test norms depends on whether the clinical trial is an efficacy or effectiveness study. For the former, test norms are not essential, but for the latter well-established norms are essential to assess meaningful change. Fourth, effect sizes may be a more meaningful measure of change because large effect sizes in improvement may only translate into small changes in percentile scores for very impaired patients (e.g., <1st percentile to 5th percentile).

5.2. Co-norming

Experts were asked about the importance of having norms from the same normative sample for the cognitive battery, whether the availability of such norms should guide test selection, whether it is possible to accurately combine information from different normative samples across measures, and whether there are any specific disadvantages of using this approach in a clinical trial. Experts felt less strongly about the importance of norms being drawn from the same sample than they did about having good norms in general. The median rating for this criterion was 6 falling in the “helpful, but not essential” classification category. Many expressed the idea that co-norming was a desirable step in test development, but that it was not necessary at this point in the process. Some experts expressed concern that drawing norms from existing independent reference groups can be misleading because the reference groups are not directly comparable. Others, however, argued that as long as the samples are sufficiently large and representative the norms need

not come from the same sample. Finally, there was a suggestion that obtaining norms for change without intervention would serve as a valuable benchmark for what constitutes meaningful change in performance with intervention.

5.3. Test administration and interpretation in diverse subgroups

Experts were asked about the importance of the availability of existing normative data for demographic subgroups in determining test selection, and to what extent the applicability of tests in multiple languages is an issue. This criterion received a median rating of 6.5. Most saw this criterion as a desirable feature for the battery, but one that could be accomplished at a later stage.

6. Cognitive domains and their integration

The first subsection addressed the importance of having adequate coverage of cognitive domains represented by the battery. For the second subsection, experts rank ordered their choice of cognitive constructs for inclusion. Sections 6.3 and 6.4 dealt with internal consistency and the usefulness of summary scores.

6.1. Coverage of cognitive domains

Experts were asked about the importance of assessing separable cognitive factors in a cognitive battery for clinical trials in schizophrenia, whether the battery should be able to differentiate patients who have different patterns of cognitive deficits, and how important it is to differentiate drug effects on different cognitive domains. Coverage of cognitive domains was seen as an essential criterion (median rating=8.0, see Fig. 6), and it raised a number of points of consideration primarily concerning the ability to differentiate distinct cognitive factors. Many experts argued that it is important to assess effects on separable cognitive constructs, but that we are not able to separate them as clearly as we would like. In a separate article in this issue (Nuechterlein et al., 2004), data are presented from a number of factor analytic studies supporting the identification of

separable cognitive constructs in schizophrenia. During the interviews, some experts noted that identification of separable constructs through factor analytic methods may be complicated by problems associated with task difficulty, factor redundancy (e.g., tests loading on more than one factor), or factor instability. Some experts suggested that data from factor analytic studies in schizophrenia suggest a general G-factor accompanied by other relatively small separable factors.

6.2. Choice of cognitive domains

Experts were asked to rank order the importance of selected cognitive domains to the treatment of schizophrenia. As shown in Table 2, executive functions, attention/vigilance, memory processes, and problem-solving ability were ranked highest. The results from the survey largely overlap with the summary of findings from the separable cognitive factors article, although the dimensions were split somewhat differently based on factor analytic results (Nuechterlein et al., 2004). Following the April meeting, the MATRICS Neurocognition Committee made a decision about cognitive targets that was based on the integration of the factor analytic results, the survey data, and discussion at the meeting. Seven cognitive domains were selected for inclusion in the battery. These included: (a) attention/vigilance, (b) speed of processing, (c) working memory, (d) verbal learning and memory, (e) visual learning and memory,

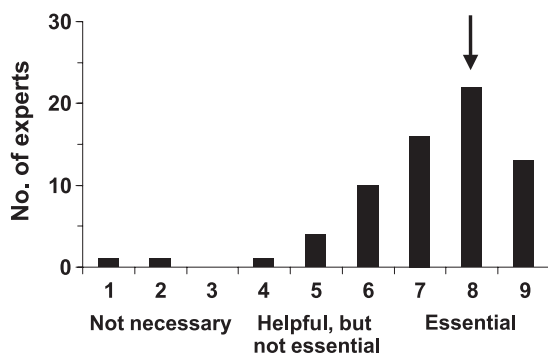


Fig. 6. Coverage of cognitive domains. How important is it for the selected battery to reliably assess the major separable cognitive factors in schizophrenia?

Table 2

Choice of cognitive domains

Test quality	Median rank	1st place votes
Executive, concept formation, and cognitive control	2	18
Attention/vigilance	3	16
Short-term memory	4	4
Long-term memory	4.5	12
Problem-solving and decision-making	5	3
Speed of processing	7	4
Social cognition	7	2
Response generation or fluency	7	1
Verbal comprehension, naming, and vocabulary	8.75	1
Early sensory and perceptual processes	9	1
Motor speed and dexterity	9.25	1
Visuospatial abilities	10	0

* Median rank: 1=most important, 9=least important.

(e) reasoning and problem-solving, and (g) social cognition.

6.3. Internal consistency

Experts were asked to what extent internal consistency should be a feature of measures to be included in the cognitive battery. This criterion received a median rating of 7.0, falling in the “essential” classification category. At the level of cognitive domain, it was viewed as necessary for tests that are grouped together to measure the same thing. However, at the level of individual tests, it was viewed as less important. One expert argued that most neurocognitive tests that will be considered for a consensus battery do not lend themselves to evaluation of internal consistency in the same way that questionnaires do. That is, most neurocognitive tests do not have item independence, a necessary characteristic for examining internal consistency.

6.4. Summary scores

Experts were asked what role summary scores should play in deciding whether a drug improves cognitive deficits in schizophrenia, and how much the psychometric properties of a summary score should matter in the selection of a cognitive test battery. Of all

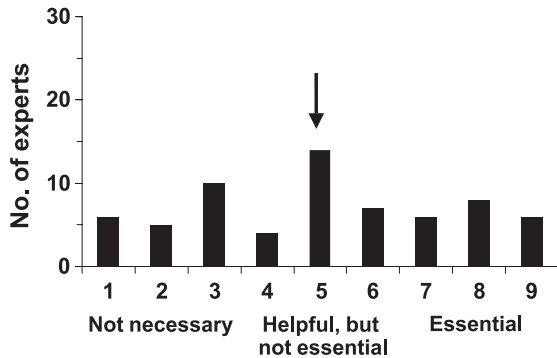


Fig. 7. Summary score. How important is it for the battery to yield a summary score that combines results across cognitive domains?

questions on the survey, this one yielded the widest range of ratings (see Fig. 7). Opinions varied widely. Arguments for a summary score included: (a) to the extent that separable factors correlate with a single factor, then a summary score is desirable; (b) a summary score is more reliable than individual domain scores; (c) the strongest correlates between cognition and functional outcome occur with cognitive summary scores; and (d) a summary score is easier to convey. Arguments against a summary score include: (a) there is no cognitive theoretical construct that maps on to a summary impairment index; (b) a summary score may mask specific effects of drugs on certain cognitive domains; (c) a summary score results in loss of potentially valuable information regarding differential treatment effects on different cognitive domains; and (d) scores from a small number of individual domains are no more difficult to interpret than one overall summary score.

7. Duration of test battery

The experts were asked for the minimum, maximum, and optimal duration of the test battery for clinical trials. This question was viewed as pitting two competing goals against one another. Brief batteries are often desirable for multisite clinical trials as a way to reduce subject demands and study expense; extended batteries are more reliable, can cover multiple cognitive domains, and are more sensitive to treatment effects. A balance between the two was seen as optimal. Fig. 8 shows the median selections for minimum, maximum, and optimal duration. Minimum

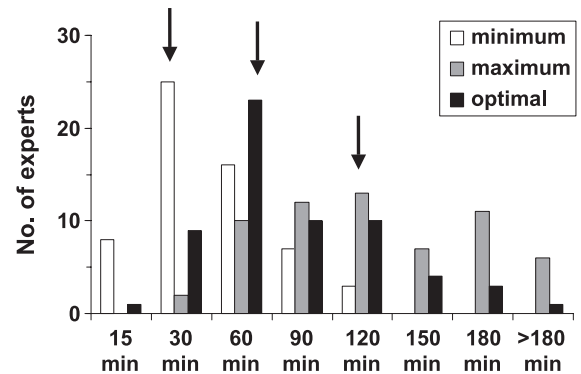


Fig. 8. Duration of battery.

duration was seen as 30 min, maximum was 120 min, and optimal was 60 min.

8. Overall ranking of importance of test qualities

At the end of the survey, experts were asked to rank order the importance of test qualities for selecting cognitive tests for the battery. As shown in Table 3, three test qualities were ranked highest, receiving a median rank of 3. These were high test–retest reliability, good coverage of key individual cognitive concepts, and comparable alternate forms. Other test qualities seen as important were: clear relationship to

Table 3
Ranking of test qualities

Test quality	Median rank	1st place votes
High immediate test–retest reliability	3	17
Good coverage of key individual cognitive concepts	3	17
Comparable alternate forms	3	1
Clear relationship to functional outcome	4	13
Strong internal consistency of individual scales	5	6
Well-established norms for general population	5	0
Highly interpretable overall summary score	7	6
Clear relationship to known neural systems	7	5
Clear relationship to clinical symptoms	8	0

* Median rank: 1=most important, 9=least.

functional outcome, strong internal consistency of individual scales, and well-established norms for the general population. The Neurocognition Committee decided on the final set of essential criteria based on survey findings, presentations by experts on psychometrics and research design at the meeting, and audience discussions at the meeting.

The final test selection criteria were: (a) test–retest reliability, (b) utility as a repeated measure, (c) relationship to functional outcome, (d) potential changeability in response to pharmacological agents, and (e) practicality and tolerability. These criteria became the basis for evaluation of candidate cognitive tests for inclusion in the final version of the consensus battery. The steps involved in this process of test selection, including the use of the RAND Panel Method, will be the subject of upcoming articles.

List of Experts

Artiola, Lidia	Iverson, Grant L.
Asarnow, Robert F.	Javitt, Daniel C.
Barch, Deanna	Keefe, Richard
Bellack, Alan S.	Krystal, John H.
Bilder, Robert M.	Levin, Edward
Block, Gilbert	Lieberman, Jeffrey A.
Braff, David	Luck, Steven J.
Breier, Alan	McCarley, Robert W.
Brekke, John	McCarthy, Gregory
Buchanan, Robert W.	McQuade, Robert D.
Callicott, Joseph H.	Meltzer, Herbert Y.
Cannon, Tyrone D.	Millis, Scott R.
Carpenter, William T.	Mintz, Jim
Carrazana, Enrique	Mohs, Richard C.
Carter, Cameron	Mueser, Kim T.
Cornblatt, Barbara	Park, Sohee
David, Tony	Patterson, Thomas L.
Essock, Susan	Penn, David
Freedman, Robert	Posner, Michael I.
Frese, Frederick	Purdon, Scot E.
Frith, Chris	Randolph, Christopher
Geyer, Mark A.	Risch, S. Craig
Gharabawi, Georges	Robbins, Trevor W.
Goff, Donald C.	Romano, Steve
Gold, Jim	Schooler, Nina R.
Goldberg, Terry E.	Seidman, Larry J.
Goldman-Rakic, Patricia	Stern, Robert A.
Gur, Raquel E.	Strauss, Milton E.
Gur, Ruben C.	Sweeney, John A.
Harvey, Philip D.	Tamminga, Carol
Heinrichs, Walter	Tulsky, David S.
Hoff, Ann L.	Velligan, Dawn I.
Hogarty, Gerard E.	White, Travis
Holzman, Philip S.	Zalcman, Steven

Acknowledgments

An earlier version of this paper was presented at: Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Conference #1: Identifying Cognitive Targets and Establishing Criteria for Test Selection, Bethesda, MD. Support for this conference came from NIMH Contract MH22006 (S.R. Marder, P.I., M.F. Green, Co-P.I.).

References

- Anastasi, A., 1988. *Psychological Testing*. (Sixth Edition). Macmillan, New York.
- Bellack, A.S., Gold, J.M., Buchanan, R.W., 1999. Cognitive rehabilitation for schizophrenia: problems, prospects, and strategies. *Schizophrenia Bulletin* 25, 257–274.
- Fitch, K., Bernstein, S.J., Aguilar, M.D., Burnand, B., LaCalle, J.R., Lazaro, P., van het Loo, M., McDonnell, J., Vader, J.P., Kahan, J.P., 2001. The RAND/UCLA Appropriateness Method User's Manual. RAND, Santa Monica, CA.
- Green, M.F., 1996. What are the functional consequences of neurocognitive deficits in schizophrenia? *American Journal of Psychiatry* 153, 321–330.
- Green, M.F., Kern, R.S., Braff, D.L., Mintz, J., 2000. Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the “right stuff”? *Schizophrenia Bulletin* 26, 119–136.
- Green, M.F., Kern, R.S., Heaton, R.K., 2004. Longitudinal relationships of cognition and functional outcome in schizophrenia: implications for MATRICS. *Schizophrenia Research*, 72, 41–51, this issue.
- Lezak, M.D., 1995. *Neuropsychological Assessment*. (Third Edition). Oxford University Press, New York.
- Nuechterlein, K.H., Asarnow, R.F., Subotnik, K.L., Fogelson, D.L., Ventura, J., Torquato, R., Dawson, M.E., 1998. Neurocognitive vulnerability factors for schizophrenia: convergence across genetic risk studies and longitudinal trait/state studies. In: Lenzenweger, M.F., Dworkin, R.H. (Eds.), *Origins and Development of Schizophrenia: Advances in Experimental Psychopathology*. American Psychological Association, Washington, DC, pp. 299–327.
- Nuechterlein, K.H., Barch, D.M., Gold, J.M., Goldberg, T.E., Green, M.F., Heaton, R.K., 2004. Identification of separable cognitive factors in schizophrenia. *Schizophrenia Research* 72, 29–39, this issue.
- Wechsler, D., 1997. *Wechsler Adult Intelligence Scale—III*. The Psychological Corporation, San Antonio, TX.