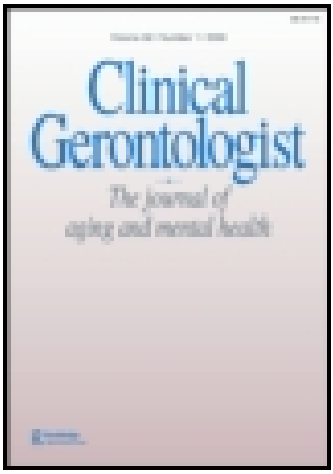

Clinical Gerontologist
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/wcli20

Utility of the Judgment Questionnaire Subtest of the Neurobehavioral Cognitive Status Examination in the Evaluation of Individuals with Alzheimer's Disease

Denise C. Woods PhD [a], Marian B. Patterson PhD [a] & Peter J. Whitehouse MD, PhD [a]

[a] University Alzheimer Center, University Hospitals of Cleveland and Case Western Reserve University, Cleveland, OH, USA

Published online: 11 Oct 2008.

To cite this article: Denise C. Woods PhD, Marian B. Patterson PhD & Peter J. Whitehouse MD, PhD (2000) Utility of the Judgment Questionnaire Subtest of the Neurobehavioral Cognitive Status Examination in the Evaluation of Individuals with Alzheimer's Disease, Clinical Gerontologist, 21:4, 49-66, DOI: 10.1300/J018v21n04_05

To link to this article: http://dx.doi.org/10.1300/J018v21n04_05

Utility of the Judgment Questionnaire Subtest of the Neurobehavioral Cognitive Status Examination in the Evaluation of Individuals with Alzheimer's Disease

Denise C. Woods, PhD
Marian B. Patterson, PhD
Peter J. Whitehouse, MD, PhD

ABSTRACT. The Judgment Questionnaire subtest (JQ) of the Neurobehavioral Cognitive Status Examination is a potential tool for examining aspects of executive function. Existing data, however, do not provide reliability or validity measures, or information on the effects of age, education, and gender among individuals over age 60. Because the screening technique of administration and scoring produces scoring anomalies and the scoring criteria are often inadequate to make scoring

Denise C. Woods, Marian B. Patterson, and Peter J. Whitehouse are affiliated with the University Alzheimer Center, University Hospitals of Cleveland and Case Western Reserve University, Cleveland, OH.

Address correspondence to: Marian B. Patterson, PhD, University Alzheimer Center, 12200 Fairhill Road, Cleveland, OH 44120.

The authors would like to acknowledge the contributions of Jon Stuckey, PhD, Case Western Reserve University, who played an important role in data management and analysis; John Bolger, PhD, Michelle Jacobs, MA, Jeff Mixon and Beth Warshawsky, who assisted with the inter-judge reliability study; and the staff of the Case Western Reserve University/University Hospitals of Cleveland Alzheimer's Disease Research Center who assisted in recruitment and diagnosis of research participants, data collection, and data management.

This study was supported in part by grant AG08012 from the National Institute of Aging, grant NH4344-01 from the National Institute of Mental Health, and an Alzheimer Center grant from the Ohio Department of Aging.

decisions, we broadened and clarified scoring criteria, and administered and scored all four JQ items. To evaluate the utility of the JQ, we examined its reliability, its discriminant and construct validity, and the effects of age, gender, and education in healthy elderly volunteers and individuals with probable Alzheimer's disease. *[Article copies available for a fee from The Haworth Document Delivery Service: 1-800-342-9678. E-mail address: <getinfo@haworthpressinc.com> Website: <http://www.HaworthPress.com>]*

KEYWORDS. Judgment, executive function, Alzheimer's disease, neuropsychology

Executive skills are those skills that allow individuals to execute purposeful, goal-directed behavior. Luria (1973) described four steps involved in the execution of a task: create an intention, form plans and programs of action, inspect performance and regulate behavior so it conforms to those plans and programs, and verify that the effects of the actions are consistent with the original intentions and correct mistakes. Executive functions are often detrimentally affected by damage to the frontal lobes, but also appear especially vulnerable to the effects of generalized cerebral injury and disease processes, such as the dementias, even in the absence of other cognitive deficits (Lezak, 1983), perhaps because frontal lobe activity is especially dependent on input from other cortical and subcortical projections (Duncan, 1986).

Because they affect regulation of all behaviors (Lezak, 1983), impaired executive skills are likely to have a pervasive, detrimental effect on autonomous function and decision making abilities. Though impaired executive skills are apparent in the course of daily living (Bonner, Cobb, Sweet and White, 1953; Lezak, 1978a), disorders of executive function can elude discovery in typical diagnostic settings, where structured interviews and neuropsychological tests directed by the examiner restrict the opportunity to observe an individual's behavior, and routine mental status exams overlook executive functions altogether (Royall, Mahurin, True, Anderson, Brock, Freeburger, and Miller, 1983). One way to probe for disturbances of executive function is to pose complex problems and observe the subject analyzing the components of the problem and formulating a strategy for the solution, an approach Luria considered to be the ''most delicate test for the diagnosis of frontal lobe lesions''(Luria, 1973).

One potential tool for examining some aspects of executive skills is the Judgment Questionnaire, (JQ) a subtest of the Neurobehavioral Cognitive Status Examination (NCSE; the Northern California Behavioral Group, 1988), designed to examine judgment by posing four problematic situations and asking the person to describe what he or she would do in response. One of the advantages of the questionnaire is its apparent focus on executive functions and its relatively small reliance on other cognitive skills. Since the respondent need not actually carry out the activities described in the response, the test requires minimal attentional, perceptual, motor, and memory abilities. Language skills only sufficient to comprehend the questions and communicate responses are needed. Of primary relevance are a person's ability to identify the problem posed in the question, formulate a hypothetical strategy for coping with the situation effectively, anticipate the outcome of the plan, and reject inadequate or inappropriate solutions.

These qualities make the test easy to administer even to individuals with sensory or motor impairments. It can be administered fairly quickly and can easily be incorporated into bedside interviews. The questions are straightforward, concrete, and easy to understand, and their practical quality often elicits cooperation from people who are impatient with more abstract tests. The literature indicates the NCSE is being employed to assess cognitive status in various psychiatric, geriatric, medical, and rehabilitation settings (Logue, Tupler, D'Amico and Schmitt, 1993; Lamarre and Patten, 1994; Mitrushina, Abara and Blumenfeld, 1994; Fields, Fulop, Sachs, Strain, and Fillit, 1992).

Problems exist, however, in the application and interpretation of the JQ. The manual offers minimal guidelines to assist the examiner in administration and scoring. It also provides little data to support interpretation of the test results, such as the reliability and validity of the subtests, and the effects of such variables as age, education, gender, and race. The existing normative data are sparse. The manual of the published test reports data are from young and old normal participants, and from geriatric and neurosurgical participants. Among the 11 subtests in the NCSE, the JQ had the least ability to identify individuals with neurosurgical diagnoses. Individuals with neurosurgical problems, however, are likely to have focal lesions, which, unless located in the frontal lobes, are less likely to affect executive functions than are more diffuse pathological processes. Thus, its sensitivity to pa-

tients with more generalized impairments, such as Alzheimer's disease, who frequently show executive impairments, (Patterson, Mack, Geldmacher & Whitehouse, 1996), has not yet been demonstrated. Though the NCSE has been examined within geriatric settings (Fields, Fulop, Sachs et al., 1992) the JQ has not been studied individually.

In the present study we evaluated the utility of the JQ in research participants with suspected Alzheimer's Disease, whose often subtle impairments of judgment may be revealed by this test. Individuals with Alzheimer's disease were thought to make a good testing ground for the evaluation of the test because the pervasive nature of the disease allows the comparison of individuals with a broad range of cognitive weaknesses, and their often peculiar responses present a challenge for administration and scoring of the test. Accordingly, in a sample of mildly to moderately impaired AD patients and normal elderly participants, we evaluated the JQ with respect to its reliability, sensitivity to demographic variations, discriminant and construct validity.

METHODS

Research Participants

Participants with Alzheimer's disease were 95 individuals enrolled in the research registry of the University Hospitals of the Cleveland Alzheimer Center, with a diagnosis of probable AD, made according to criteria specified by the National Institute of Neurological and Communicative Diseases and Stroke-Alzheimer's Disease and Related Disorders Association work group, the Health and Human and Services Task Force on Alzheimer's Disease (McKhann, Drachman, Folstein, Katzman, Price, and Stadlan, 1984). These criteria include evidence of progressive decline in functioning and deficits in two or more areas of cognition. Exclusion criteria included symptoms of other disorders that could account for the decline and deficits; a history of alcoholism or substance abuse within five years of the onset of dementia; evidence of Parkinson's disease requiring therapy; focal central neurologic illness, signs, or symptoms; serious metabolic or toxic disorders; serious cardiac disease; sensory loss sufficient to affect testing; current major depression or comparable psychotic illness.

Informed consent was obtained from each participant and/or a family member, when appropriate.

The comparison group consisted of 49 research volunteers enrolled in the research registry. They were recruited from spouses of individuals in the AD registry, hospital volunteers, and individuals within the same community as the AD registry. They underwent the same medical and neuropsychological evaluations and were subject to the same exclusion criteria as the individuals with AD but were judged free from cognitive impairment by a PhD neuropsychologist. Written informed consent was obtained from each individual.

Materials and Procedure

In addition to the JQ, the following measures were used in the present study: Two measures of generalized impairment: Total Score of the Mini Mental State Exam (MMSE) (Folstein, Folstein & McHugh, 1975); and Total Score of a version of the Blessed Dementia Scale, (DS), used in the Consortium to Establish a Registry for Alzheimer's Disease (Heyman, 1987; Morris et al., 1989) a measure based on interview with an informant regarding the patient's competence in activities of daily living; the Test Age from the Porteus Maze Test (PMT) (Porteus, 1965; Mack & Patterson, 1995), a paper and pencil maze; two measures of visual perception, the identification level of the Gollin Incomplete Pictures Test (GIPT) (Gollin, 1960; Mack et al., 1993) in which individuals are asked to identify fragmented pictures of common objects; and the Total Accuracy Score from the CERAD Constructional Praxis test, (CP), in which individuals are asked to copy four simple designs; performance on two brief langauage tests from the CERAD, Total Score of the Verbal Fluency (VF), the number of animals named in 60 seconds; and total of pictures correctly named on a 15 item version of the Boston Naming Test (BNT); two measures of verbal secondary memory obtained from the CERAD 10 word list learning task: number of words correctly recalled after a brief delay (DLY), and number of words recognized from a list of 10 targets and 10 distractors (RCG); a measure of attention, total number of sequences correctly reproduced in the Nonverbal Sequencing Span, (NSS), a modified Knox Cubes test (Arthur, 1947, Winegardner & Mack, 1992) in which subjects are asked to repeat a sequence of taps on four blocks.

The JQ consists of four questions concerning what one would do in

the following situations: (1) being stranded out of town in an airport with no resources other than a dollar, (2) waking up one minute before an important appointment on the other side of town, (3) finding water pouring into the kitchen from a broken pipe, and (4) seeing a child, age two, unsupervised, at the end of a pier.

The JQ is administered according to a screening metric: the most difficult item is administered first and, if passed, is the only item administered, and the respondent is awarded 5 points; if failed, the other three items are administered and scored on a scale of 0 to 2 and totaled to create a score that can range from 0 to 6. The peculiarity exists, therefore, that respondents who do as well as possible do not earn as many points as some respondents who fail the first, screening item. To avoid this incongruous situation, we administered all four questions to each subject. In addition, administering all the items broadened the opportunity to observe qualitative aspects of the individual's executive functions. The items were administered to all participants in order of difficulty: 2, 3, 4, 1. Examiners recorded the responses both in writing and on audio tape, and typed transcriptions of the tape were checked by the examiner for accuracy.

In our clinical experience with the test, we found examiners frequently were uncertain how to score the responses. We, therefore, developed guidelines to facilitate the scoring of responses that did not cleanly fit the existing limited criteria. In addition, criteria were developed for one point responses to item 1, which in the published version of the test had only a two and a zero point scoring system.

Inter-judge reliability was measured in a subset of the data of 11 protocols of 9 individuals without AD and those of 20 individuals with AD. These responses were typed and presented to the judges with no indication of the diagnostic status of the individual. The judges were chosen from a range of educational and experience backgrounds: a PhD neuropsychology fellow, an MA neuropsychology assistant, a BA psychology research assistant, and a BFA research assistant. All were experienced in administering neuropsychological tests, but only the MA neuropsychology assistant had direct experience with the JQ prior to this study. The judges were instructed to score the responses according to the written instructions (i.e., those presented in the manual supplemented by our expanded guidelines) but were given no further training or discussion, in an effort to replicate the typical clinical situation in which the examiner administers and scores the test solely

on the basis of information provided in the test manual. Scores for each item and a total score for each protocol were obtained from each judge.

RESULTS

Inter-Judge Reliability

The inter-judge reliability for each of the four items was obtained using interclass correlations. They are for item 1, .79; for item 2, .89; for item 3, .91; and for item 4,.75; and for the JQ Total score inter-judge reliability was .69.

Internal Consistency

The characteristics of the Comparison and AD group for the internal consistency and correlational analyses are listed in Table 1. Not all participants were included in each analysis, since not all data were available from each participant.

Inter-item correlations among the four JQ items were generally low,

TABLE 1. Demographic Characteristics of Comparison and AD Groups Within Each Statistical Analysis

	N	Age	Education	Gender
Subjects used for analysis of internal consistency of the JQ				
Comparison group	49	69.3 (6.3)	15.5 (3.2)	M = 21, F = 28
AD group	81	72.7 (7.3)	12.8 (3.1)	M = 30, F = 51
Subjects used for correlational studies between JQ and the demographic, and neuropsychological measures				
Comparison group	48	69.1 (6.3)	15.4 (3.2)	M = 20, F = 28
AD group	79	72.7 (7.3)	12.8 (3.0)	M = 28, F = 51
Subjects used for comparison of total JQ scores in normal and AD groups				
Comparison group	22	67.5 (5.7)	15.0 (3.0)	M = 10, F = 12
All AD	38	73.7 (6.9)	12.8 (3.2)	M = 12, F = 26
High functioning AD group (ADH)	15	72.8 (6.2)	12.9 (2.5)	M = 3, F = 12
Moderate functioning AD group (ADM)	22	74.3 (1.5)	12.8 (3.6)	M = 9, F = 14

ranging from negative .01 to .38 for the AD group, and from negative .03 to .26 for the Comparison group. The internal consistency of the JQ, based on the standardized alpha statistic, was .04 for the 49 subjects in the Comparison group and .46 for the 81 participants with AD. In Table 2 is shown the total score variance when a given item is removed. The item that contributed the greatest variation to the total score among the Comparison group was number 1, the most difficult item; and among the AD group was, number 2, the easiest item.

Correlations of Demographic Variables with Test Scores

The correlation of demographic variables with the total score and the individual items is presented in Table 3, along with the number of subjects included in each correlation. The number of scorable re-

TABLE 2. Total Scale Variance If Item Is Deleted for Comparison Group and AD Group

Item	1	2	3	4
Comparison group	.87	1.22	.92	1.04
AD group	2.36	1.84	2.19	2.21

TABLE 3. Correlation of JQ Items and Total with Demographic Variables

	Item #1	Item #2	Item #3	Item #4	Total
Comparison group					
N's	38	38	34	46	22
Age	0.02	0.07	−0.16	−0.11	−0.01
Education	−0.09	−0.07	0.03	−0.3	−0.18
Sex†	0.03	−0.05	−0.37	−0.01	−0.31
AD group					
N's	54	50	50	58	28
Age	0.1	−0.02	0	−0.27	−0.02
Education	0.17	−0.02	0.31	−0.02	0.26
Sex†	0.01	0.25	−0.40*	−0.1	−0.05

*$p < .01$
**$p < .001$
†Negative correlations with sex variable indicate better performance among males.

sponses available varied for each item and for the total. Because of the large number of correlations generated in this table, a strict alpha, .01, was employed to determine significance. Age did not correlate significantly with any JQ item or the Total among either the Comparison group or the AD group. Education correlated with JQ significantly only among the AD group and only with Item 3: $r = .31$, $p < .01$. Gender correlated only with Item 3: *all* males in the comparison group gave the full two point response, and only half the females did so; $r = -.37$, $p < .01$. Among individuals with AD, the results only approached significance: half the males gave the full two point response, compared with only about 12% of the females: $r = -.40$, $p < .03$. Because our sample only included three non-whites, the effects of race could not be assessed.

Criterion Validity

The criterion validity of the total JQ score was first tested by comparing the performance of AD and Comparison groups. Demographic data for the groups used in the Total score comparison are shown in Table 1. The groups differed with respect to age ($t = -3.54$, $df = 58$, $p < .002$) and education ($t = 2.58$, $df = 57$, $p < .02$), but not gender. Scores for the individual JQ items and Total score for Comparison and AD groups are presented in Table 4. The number of subjects with Total JQ scores was dependent on a given subject having produced scorable responses for all four JQ items, and was thus smaller than the number of subjects available for analysis of individual items.

TABLE 4. Judgment Questionnaire Scores

	Item #1	Item #2	Item #3	Item #4	Total
Comparison Group					
N	34	46	38	38	22
Mean (s.d.)	1.29 (.76)	1.94 (.25)	1.76 (.49)	1.66 (.58)	6.50 (1.23)
Range	0-2	1-2	0-2	0-2	3-8
AD Group					
N	62	73	68	64	38
Mean (s.d.)	.57 (.76)	1.58 (.74)	1.06 (.75)	1.61 (.70)	4.82 (1.92)
Range	0-2	0-2	0-2	0-2	0-8

The mean Total score for the Comparison group, 6.5 (range = 3-8, sd = 1.2, n = 38), differed significantly from that for the AD group, 4.8 (range = 0-8, sd = 1.9, n = 22), $F(1,58) = 13.71$, $p < .0005$.

A hit rate analysis was also performed. Since the JQ only assesses judgment and thus is not a diagnostic test for AD, the results of the analysis must be interpreted cautiously. With the cutoff score set at four and below for AD, the specificity was 95%, the sensitivity was 42%, and the proportion correctly classified was 62%. Because greater sensitivity is preferred in some settings, an analysis was also performed applying a cutoff of 5 and below. The resulting specificity was 86%, sensitivity was 61%, and the proportion correctly classified was 70%.

The diagnostic value of the original screening metric (The Northern California Behavioral Group, 1988) was also examined. The mean and standard deviation for the Comparison group were 5.1, (identical to the normative data) and .64, and for the AD group were 4.3 (1.5). When compared in a one way ANOVA, $F(1,67) = 7.43$, $p < .008$. In a hit rate analysis using the screening metric with a cutoff score set at three and below to predict impaired judgment, the specificity rate was 95%, the sensitivity rate was 32%, and the proportion correctly classified was 50%.

Because diagnostic questions arise most frequently in the early stages of AD, we next divided the AD group into high functioning (N = 15) and moderately to severely impaired (N = 23) groups of individuals with AD (ADH, ADM), as defined by a score of 20 or higher on the Mini Mental Status Exam, or lower than 20, respectively. Demographic characteristics of the ADH and ADM groups are shown in Table 1. The AD groups did not differ from each other with respect to age, education, or gender, although both differed significantly from the Comparison group with respect to age, as demonstrated with a one way ANOVA comparing all three groups on age ($F(2,57) = 6.42$, $p < .004$), followed by post hoc Newman-Keuls step-wise analysis. The mean JQ Total Score for the ADH group was 5.5 (sd = 1.2) and for the ADM group, 4.3 (sd = 2.2). A one way ANOVA comparing all three groups (ADH, ADM and Comparison group) produced an $F(2,57) = 9.65$, $p < .0002$. A post hoc Newman-Keuls step-wise analysis, however, showed that only the ADM group differed significantly from the Comparison group and the ADH group, who did not differ significantly from each other.

Finally we examined correlations within the AD group between each item of the JQ and the total score with measures of general severity of impairment, as indexed by the MMSE, and by a measure of functional ability, the DS. Pearson *r* correlations of the JQ measures with the two severity measures are presented in the first two rows of Table 5. Because of the large number of correlations, a strict definition of significance was employed: alpha was set at .01. Only Item 2 produced significant correlations, and with both the DS and the MMSE. No correlation is reported between the measures of general severity and the JQ among the CG, because they showed essentially no variation on the measures of general dementia severity, (MMSE mean = 29.25, sd.= .86, Range = 27-30; DS, mean = .06, sd. = .22, Range = 0-1.0).

Construct Validity

Correlations between the JQ measures and the eight neuropsychological measures were carried out separately for the Comparison group and the AD group and are presented in the last eight rows of Table 5.

Convergent Validity. Convergent validity was estimated by examining the correlations between the JQ and another test sensitive to deficits of executive functioning, the PM, none of which was found to be significant.

Divergent Validity. Divergent validity was assessed by the correlations between the JQ and the GIPT and the CP, measures of perception and design copy, cognitive skills not expected to contribute to verbal reasoning. Within the Comparison group, no significant correlation was obtained. Within the AD group, CP correlated significantly with the Total score.

To confirm our expectations that memory has negligible influence over JQ performance, we examined correlations between the JQ and the two measures of memory. Within the Comparison group, only RCG showed a significant correlation, with Item 3. Within the AD group, no correlation was significant.

To confirm our expectations that attention contributed little to JQ performance, we examined correlations between the JQ and a measure of attention. Within the Comparison group, NSS correlated significantly (and negatively) only with Item 4. Within the AD group, no correlation was significant.

TABLE 5. Correlation of JQ Items and Total with Other Neuropsychological Tests

	Judgment Questionnaire Item									
	1		2		3		4		Total	
Blessed DS	*r*	N	*r*	N	*r*	N	*r*	N	*r*	N
AD group	−0.23	(45)	−0.33*	(51)	−0.30	(47)	−0.25	(44)	−0.41	(26)
MMSE										
AD group	0.26	(50)	0.39*	(58)	0.25	(54)	0.01	(50)	0.34	(28)
Porteus Mazes										
Comparison group	0.07	(32)	−0.13	(43)	0.12	(36)	0.09	(35)	0.13	(21)
AD group	0.19	(44)	0.07	(52)	0.19	(47)	0.21	(43)	0.28	(23)
Gollin										
Comparison group	−0.08	(24)	−0.15	(34)	−0.08	(26)	0.31	(28)	0.18	(13)
AD group	0.00	(47)	−0.11	(54)	−0.17	(50)	0.14	(47)	−0.23	(26)
Construct'l Praxis										
Comparison group	−0.24	(34)	−0.02	(46)	0.16	(38)	0.39	(38)	0.02	(22)
AD group	0.21	(50)	0.28	(58)	0.32	(54)	0.18	(50)	0.53*	(28)
Verbal Fluency										
Comparison group	−0.08	(34)	−0.22	(46)	−0.22	(38)	−0.17	(38)	−0.13	(22)
AD group	0.17	(50)	0.50**	(58)	0.44**	(54)	0.13	(50)	0.66**	(28)
Boston Naming										
Comparison group	−0.14	(34)	0.10	(46)	−0.04	(38)	−0.14	(38)	0.07	(22)
AD group	0.17	(50)	0.27	(58)	0.40*	(54)	0.27	(50)	0.57*	(28)
Word Recall										
Comparison group	−0.06	(34)	−0.24	(46)	0.05	(38)	0.03	(38)	−0.23	(22)
AD group	0.02	(50)	0.05	(58)	0.30	(54)	−0.16	(50)	0.27	(28)
Word Recognition										
Comparison group	−0.31	(34)	−0.15	(46)	0.43*	(38)	−0.13	(38)	−0.24	(22)
AD group	−0.02	(50)	0.25	(58)	0.21	(54)	0.24	(50)	0.21	(28)
Nonverbal Seq.										
Comparison group	0.03	(32)	−0.05	(45)	0.02	(36)	−0.47*	(36)	−0.23	(21)
AD group	0.24	(50)	0.02	(58)	−0.04	(54)	−0.05	(50)	0.06	(28)

*$p < .005$
**$p < .001$

Influences of Verbal Skills. Since AD often causes verbal deficits, we examined the influence of verbal abilities on the JQ. Correlations between the JQ and the verbal measures are shown in Table 5. Within the AD group, significant correlations were seen between the Total score and VF (r = .66) and BN (r = .57); between Item 2 and VF (r = .54) and BN (r = .40) and between Item 3 and VF (r = .44).

DISCUSSION

Inter-judge reliability was generally acceptable. It should be remembered that we provided our raters with guidelines concerning scoring criteria, but they received no additional training on their use. With more training, reliability would likely improve.

The low internal consistency may reflect in part the small number of items of the test, however it also suggests that the test is not narrowly focused on a single construct. Judgment is, in fact, a complex concept involving several aspects of executive function that may not covary, and that may differentially influence the responses to each item. For instance, judgment comes into play as an individual attempts to grasp the problem posed by the situation described in the question. Inspection of the poor responses suggests that many individuals aim their concerns in peculiar directions or become distracted by tangential aspects of the question. Judgment also influences how well the person formulates an appropriate solution to a specific problem, an ability likely to be influenced by an individual's personality. Self-confidence, assertiveness, and gender thus may influence the degree to which individuals are self-reliant or accustomed to look to or impose on others for help. Consideration of and a sense of responsibility toward others may also be a factor, since two of the questions pose situations which hypothetically affect primarily the examinee, while the other two primarily affect someone else.

In addition to the examinee variables, the questions themselves pose different types of problems. While this property is generally desirable to avoid focusing on specialized skills, these differences may reduce internal consistency. For example, Item 3 is clearly dependent on particular knowledge: when a broken pipe is flooding the kitchen, one turns off the main water valve. Among the men and half the women in our comparison group, that information appears to be well known. The question becomes for them, then, not a test of judgment,

but a test of knowledge. For the respondents who do not know the correct answer, the question becomes a test of judgment, but the range becomes truncated to one or zero points. Items 1, 2, and 4 are heavily dependent on social judgment: what are one's social obligations when one will be late to an appointment, what are one's obligations toward the children of strangers, what are the responsibilities of an airline toward its customers in distress, etc.

On the basis of the correlations of demographic variables with JQ scores, it appears that the JQ is fairly robust with regard to age, gender, and educational variables, with some noted exceptions, especially item #3. Sensitivity to racial and cultural influences, however, were not examined.

With respect to the sensitivity and specificity of the JQ in identifying AD, although AD appears to have a detrimental effect on the performance of the JQ, the JQ has only fair ability to discriminate between individuals with and without AD, whose scores overlapped extensively. The expanded scoring system, which includes full administration and scoring of all four items, modestly increases the test's ability to discriminate normal older individuals from those with Alzheimer's Disease; the specificity remains very high with the new system, and the sensitivity improves, though it remains poor. It should be noted that, while participants with AD were significantly older and had significantly less education than those in the Comparison group, these differences are not likely to have biased the results. First, correlations of JQ performance with age or education were rather small and generally not significant. Second, the age and education differences would be expected, if anything, to spuriously enhance the separation between the groups and thus would not alter our interpretation that the JQ is a relatively weak discriminator.

The results concerning differences between individuals with AD and normal elderly are what would be expected, however, since the JQ is not expected to serve as a marker for AD, but to reflect the variability of judgment skills among individuals with AD. Though AD is generally considered a disease with pervasive cortical effects, the lesions are not uniformly distributed throughout the cortex, resulting in wide individual variation among all aspects of cognitive skills. No one test, nor even a collection of tests reflecting abilities in any one cognitive realm, will adequately measure the decline of function resulting from AD. Consequently, while very poor scores on any test of

cognitive function nearly always indicate cognitive impairments, some individuals with AD will not exhibit those impairments on a particular task. It is this variability of impairments that may account for the substantial behavioral differences frequently observed clinically among individuals with AD.

The focus of the test on executive functions was not demonstrated. Correlations suggest that performance on the JQ is no more related to the executive abilities tapped by the Porteus Mazes than to other cognitive abilities. Other investigators (Shallice and Burgess, 1991) have found that executive ability is not a unitary construct that loads onto a single factor. To the extent the Porteus Mazes, a test of planning ability, and the JQ, a test of social problem solving, represent different aspects of executive ability, perhaps it is not surprising, in retrospect, that the scores on these two tests do not correlate highly among individuals with AD. One may also speculate that the sensitivity of the test to executive dysfunction is limited by the essentially passive nature of the test, which does not probe the respondent's actual execution of the plan stated in the response, in contrast to other measures of executive function which evaluate the respondent's actions and results.

In general, the results of these studies indicate the JQ does not serve as a specific measure of executive function, but may tap into some aspects of general cognitive function. While the correlations of JQ Total score with the MMSE (.34) and DS ($-.41$) approached moderate size, they were not significant; analysis of a larger sample size would be needed before concluding that the JQ is unrelated to generalized cognitive or functional deficits. A large portion of the JQ variance appears to be independent of the DS and MMSE, however. In view of the limited reliability of the JQ, it may not be possible to determine how much of that variance is specific to the construct of judgment. The significant correlation (.53) between JQ Total score and Constructional Praxis performance suggests no obvious interpretation and awaits further explorations.

There is evidence verbal skills, at least those involving expressive language ability, may have some influence on JQ performance. Though not tested directly in the present study, it is possible that subtle aspects of some of the items might be grasped better by those with better language comprehension skills. For example, some people seem

to ignore the information that the airport in which they are stranded is out of town.

A review of individual protocols suggested the construction of some items may produce difficulty with administration and scoring. For example, Item 1 (airport) often generates consternation among individuals who cannot entertain the hypothetical situation posed. Many cannot conceive of being without means of self-reliance unless they had been robbed, and the introduction of that assumption naturally alters the optimal response. Item 4 (pier) generated the most frequent difficulties in administration and scoring. One difficulty is that individuals sometimes seem to fail to acknowledge the inherent danger present in the situation, and some are more concerned about the child's apparent abandonment than about its risk of falling. Many seem to assume the pier is protected by a railing, which is, of course, possible, and not clarified in the question. Another possible source of difficulty is that some individuals seemed to consider themselves incapable of physically handling a child of that size, who is likely to resist a stranger, especially since our respondents are elderly.

The results of the present analyses do not encourage the use of the JQ score as an indicator of executive dysfunction in clinical diagnostic or research settings, although the quality of specific responses can be quite informative for clinical purposes. While the JQ as currently constructed has some utility, however, we believe its value could be enhanced by modifying some of the items to clarify the situation posed, eliminating the item that relies on a base of knowledge strongly influenced by gender, and perhaps improving the reliability by lengthening the scale.

REFERENCES

Arthur, G., *Point scale of performance tests*, New York: Psychological Corp., 1947.

Bonner, F., Cobb, S., Sweet, W. H., & White, J.C., Frontal lobe surgery, *Psychiatric Treatment. Proceedings of the Association for Research in Nervous and Mental Disease*, 1953, 31, 392-421.

Duncan, J., Disorganization of behavior after frontal lobe damage, *Cognitive Neuropsychology*, 1986, 3, 271-290.

Fields, S. D., Fulop, G., Sachs, C. J., Strain, J., Fillit, H., Usefulness of the Neurobehavioral Cognitive Status Examination in the hospitalized elderly, *International Psychogeriatrics*, 1992, 4, 93-102.

Folstein, M. F., Folstein, S. E., McHugh, P.R., Mini-Mental State Examination, *Journal of Psychiatric Research*, 1975, 12,, 189-198.

Gollin E. S., Developmental studies of visual recognition of incomplete objects, *Perceptual and Motor Skills*, 1960, 11, 289-298.

Heyman, A., *Consortium to Establish a Registry of Alzheimer's Disease*. Durham, NC, Duke University Medical Center, 1987.

Lamarre, C. J., Patten, S. B., A clinical evaluation of the Neurobehavioral Cognitive Status Examination in a general psychiatric inpatient population, *Journal of Psychiatric Neuroscience*, 1994, 19, 103-108.

Lezak, M. D., Living with the characterologically altered brain injured patient. *Journal of Clinical Psychiatry*, 1978a, 39, 592-598.

Lezak, M. D., *Neuropsychological Assessment*, New York: Oxford University Press, Inc., 1983.

Logue, P. E., Tupler, L. A., D'Amico, C., Schmitt, F., The neurobehavioral cognitive status examination: Psychometric properties in use with psychiatric inpatients, *Journal of Clinical Psychology*, 1993, 49, 80-89.

Luria, A. R., Translated by Basil Haigh, *The Working Brain: An Introduction to Neuropsychology*, London: Penguin, 1973, p. 220.

Mack, J. L., Patterson, M. B., Executive dysfunction and Alzheimer's Disease: Performance on a test of planning ability, the Porteus Maze Test, 1995, *23rd Meeting of the International Neuropsychological Society*, February 8-11, (10th), 1995, Seattle, Washington.

Mack, J. L., Patterson, M. B., Schnell, A. H., Whitehouse, P. J., Performance of subjects with probable Alzheimer Disease and normal elderly controls on the Gollin Incomplete Pictures Test, *Perceptual and Motor Skills*, 1993, 77, 951-969.

McKhann, G., Drachman, D., Folstein, M., Katzman, R. Price, D., Stadlan, E. M., Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's Disease, *Neurology*, 1984, 34, 939-944.

Mitrushina, M., Abara, J., Blumenfeld, A., Aspects of validity and reliability of the Neurobehavioral Cognitive Status Examination in assessment of psychiatric patients, *Journal of Psychiatric Research*, 1994, 28, 85-95.

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., vanBelle, G., Fillenbaum, G., Mellits, E. D., Clark, C., and the CERAD investigators., The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part 1. Clinical and neuropsychological assessment of Alzheimer's Disease. *Neurology*, 1989, 39, 1159-1165.

The Northern California Neurobehavioral Group, Inc., *The Neurobehavioral Cognitive Status Examination*, Fairfax, California, 1988.

Patterson, M. B., Mack, J. L., Geldmacher, D. S., & Whitehouse, P. J., Exeutive functions and Alzheimer's disease: problems and prospects, *European Journal of Neurology*, 1996, 3, 5-15.

Porteus, S. D., *Porteus Maze Test, Fifty Years' Application*, Palo Alto, Pacific Books, 1965.

Royall, D. R., Mahurin, R. K., True, J. E., Anderson, B., Brock, I. P., Freeburger, L., Miller, A., Executive impairment among the functionally dependent: Comparisons between schizophrenic and elderly subjects, *American Journal of Psychiatry*, 1983, 150, 1813-1819.

Shallice, T. and Burgess, P., (1991). Higher-order cognitive impairments and frontal lobe lesions in man. In *The frontal lobes revisited* (Ed E Perecman), pp. 125-138. IRBN Press, New York.

Winegardner, J., Mack, J. S., Sequencing span deficits in dementia, presented at the *International Neuropsycholocal Society European Conference*. Deauville, France, 1992.

RECEIVED: 08/21/99
ACCEPTED: 01/26/00