Validity of Short Forms of the Category Test

© Gwenyth L. Sterns 1987

A Thesis Submitted in Partial Fulfillment of Requirements

for a Degree in Master of Science to

Saint Mary's University

Halifax, Nova Scotia

Approved: ______________________
Irmingard Lenzer, Ph. D.
Faculty Advisor

Approved: ______________________
Philip Street, Ph. D.
Thesis Committee Member

Approved: ______________________
Ronald Stewart, M. Sc.
Thesis Committee Member

Approved: ______________________
Brian Doan, Ph. D.
Thesis Committee Member

Date: June 26, 1987

ISBN 0-315-40291-1

ACKNOWLEDGEMENTS

I wish to thank Dr. Irmingard Lenzer, Thesis Committee Chair, and Committee Members Dr. Phil Street, Ron Stewart, and Dr. Brian Doan for their helpful criticisms and advice during the preparation of this manuscript. I am grateful to Dr. Doan in his position as Head of the Psychology Department, Victoria General Hospital for allowing access to patient files and for his tolerant support throughout the course of this project. Dr. Joe Gabriel's willingness to serve as outside reader is greatly appreciated. I also thank Ms. Norma Wadden, fellow student, for her patient assistance with SPSS-X.

TABLE OF CONTENTS

TABLES

ABSTRACT

Validity of Short Forms of the Category Test

Gwenyth L. Sterns

June 26, 1987

The criterion - related validity of two short forms of the Category Test was examined. As was the case in earlier studies, both short forms yielded high validity coefficients and high agreement with respect to Halstead's cut - off score for impairment. The 120 - item form was the stronger predictor, accounting for 90.8% of the variance in the original Category Test, while the 108 - item form accounted for 79.5%. However, both short forms were seriously limited with respect to how closely the estimated scores approximated the actual scores; even the best short form prediction equations produced estimated scores within a 10 - point range of the actual scores only 44% and 42% of the time. Since the main clinical purpose for using a short form would be to obtain an estimate of the actual score, neither short form can be recommended for clinical use. This study illustrates the need to consider the specific purpose for which a predictor is to be used when investigating the criterion - related validity of a measure.

INTRODUCTION

The Halstead - Reitan Neuropsychological Test Battery enjoys wide popularity in the field of clinical neuropsychology.. Numerous studies by Reitan and others have demonstrated that this battery is as effective or more effective than most neurodiagnostic techniques (Vega & Parsons, 1967; Snow, 1981). One of the major criticisms of the battery is the long administration time, a minimum of four to six or more hours in a given clinical situation. This makes the procedure very stressful for some patients, and very expensive. There have been a number of attempts to address this problem by abbreviating the battery (Golden, 1976; Erickson, Calsyn & Scheupbach,1978), but this has involved omitting the Category Test, ostensibly because of its long administration time.

The Category Test is a relatively complex test of abstraction, concept formation, and organizational ability. Its contribution to the understanding of a patient's cognitive functioning lies in its ability to identify those individuals who can recognize, and then ignore, irrelevant aspects of the stimulus material, and thereby get to the essential nature of the problem situation (Reitan, 1967). Thus, it may be considered a measure of current adaptive ability, a complex mental process related to, but somewhat different from, formal educational training or IQ. For success on this test, the individual must have the ability to note similarities and differences in stimulus material, construct hypotheses taking into account these similiarities and differences, then test and adapt these hypotheses in accordance with confirming or disconfirming feedback (Reitan, 1966). Because of the immediate feedback and the necessity for the individual to adapt to the situation as the circumstances unfold, the test has an interactive quality shared by few other tests of cognitive functions in common use. Because the Category Test has been shown to be one of the most sensitive indicators of general cerebral integrity (Finlayson, Johnson, & Reitan, 1977; Reitan, 1955, 1967), its omission from the Halstead - Reitan Battery represents the loss of valuable clinical information.

Another approach to the problem of long administration time has been to devise an abbreviated form of the standard 208 - item Category Test. Three short forms have been proposed. Gregory, Paul, & Morrison (1979) devised a 120 - item form (hereinafter called C120) by dropping items from the second to fifth subtests and omitting the last two subtests. This approach was based on the assumption that dropping items at the end of a subtest, or completely eliminating later subtests, could not affect performance on preceeding items. Also, the effect of the loss of experience of a few items at the end of a subtest would likely be negligible since the concept changes dramatically from one subtest to the next. Support for the latter assumption was offered by Sherrill (1985), who compared the frequency of correct responses across each subtest for two samples, one which received the C120 and another whose original test protocols were rescored according to the C120 formula. The highly similar patterns of frequency of correct responses on subtests IV and V for the two samples suggested that the deletion of items at the end of subtests III and IV did not affect the process of learning a subsequent principle. Calsyn, O'Leary, & Chaney (1980) examined the utility of using only the first four subtests in their short form, which reduced the number of items to 108 (C108). This approach was felt to have the advantage of not taking items out of sequence and eliminating subtests V and VI which had been previously shown to have limited discriminative power (Boyle, 1975, cited in Calsyn et al.). Sherrill also described a third short form of 95 items, derived by reducing the first two subtests considerably and omitting the last two; this form has been in use for many years at the University of Wisconsin and T - score norms are available.

The substitution of an abbreviated form of a test widely recognized for its clinical utility is justifiable if: (a) the short form predicts the score on the original form with an acceptable degree of accuracy; that is, demonstrates high criterion - related validity; and (b) also results in a practical saving of time (Anastasi,1982, p 142). One measure of criterion - related (concurrent) validity for an abbreviated form of a currently available test is the correlation , or validity coefficient (Anastasi, 1982, p142). This coefficient indicates the degree of relationship between the predictor and criterion scores. Previous derivation and cross - validation studies of the 120 - item and the 108 - item forms of the Category Test have reported consistently high correlations (from .83 to .95) between the respective short form and the standard form of the test for a variety of patient and normal samples (Calsyn et al.,1980; Golden et al., 1981; Taylor, Goldman, Leavitt, & Kleinmann, 1984; Sherrill, 1985).

However, for the clinician whose data - based decisions may influence the life plans of the individual, the precision of the predictor is as important as the degree of relationship between the predictor test and the criterion test. Therefore, as an alternative to the Pearson correlation coefficient, the prediction problem can be considered a task in predicting specific criterion outcomes (Ghiselli, Campbell, & Zedeck, 1982, p 307). In this context, validity, or accuracy, can be described as the proportion of correct predictions. Assuming that the univariate linear model provides the best description of the relationship between the actual test and the respective short form, the prediction task may be addressed through the use of regression analysis to produce a prediction equation. By inserting the obtained (i.e., the short form) score into the equation, an estimate of the actual score can be computed and compared with the actual score in the context of a defined criterion for predictive accuracy.

The previous studies of the 120 - item and 108 - item forms of the Category Test used this procedure, defining the criterion for predictive accuracy as the degree of agreement between the estimated scores and the actual scores with respect to Halstead's cut - off score for impairment (i.e. 51 errors). For the C108, Calsyn et al. reported estimated score hit rates of 84% and 86% for their two alcoholic samples, with evenly distributed false positive and false negative errors. Of the studies that used the C108 and Calsyn et al.'s regression equation, Golden et al. reported agreement for 87 - 92% for patient samples and 100% for their normal sample, also with nearly equal false positive and false negative hit rates; and Taylor et al. found 87% agreement as to whether the number of errors was 51 or higher for their brain - damaged subjects.

Sherrill applied all three short form formulae (i.e., for the C120, the C108, and the C95) to the standard test protocols of a heterogeneous sample of out - patients referred to a private neuropsychology practice. Using the actual Category Test scores as the dependent variable and the short form scores as the independent variable, Sherrill applied linear regression procedure to obtain prediction equations and standard errors of estimate for each short form. He concluded that the C120 was the most accurate predictor because it was most strongly correlated with the original form, and had a smaller standard error of estimate than the other short forms. With regard to the standard cut - off score for impairment, in Sherrill's study the C120 equation misclassified only 4% of subjects in the false positive direction (i.e., as impaired, when on the actual form they were not), and 10% of subjects in the false negative direction.

It is essential to consider the specific purposes for which the predictor is to be used when deciding on an appropriate criterion for evaluating predictive accuracy (Ghiselli et al., p 274). While the results of previous studies indicate a high degree of one type of predictive accuracy, it can be argued that the use of the conventional cut - off score as the criterion is not sufficiently stringent for acceptance of the short forms of the Category Test for clinical use because of the limitations in the use of the cut - off score as a method of test interpretation. For example, the standard cut - off score for impairment does not allow for the effect of subject variables on level of performance. Numerous studies have demonstrated relationships between age and education with performance on the Category Test (Bak & Greene,1980; Finlayson, Johnson, & Reitan,1977; Prigatano & Parsons, 1976; Vega & Parsons, 1967), with age, in particular, having an adverse effect on test scores. Fromm - Auch and Yeudall (1983), from their review of normative studies of the Halstead - Reitan Battery, recommend that the cut - off point of 51 errors is appropriate only for those under the age of 40 years. Until age - corrected norms for the Category Test are developed, Reitan (1979) has suggested the "rule of thumb" that, beyond 60, the patient's age be used as an approximate cut - off point for impairment.

Halstead's cut - off score is just one of a number of strategies used in the interpretation of an individual's neuropsychological test data and, as such, plays a relatively minor role in arriving at inferences concerning the cognitive deficits of the individual patient. On the other hand, knowledge of where the individual's score falls on the continuum is essential for clinical interpretation of the Category Test. In most clinical situations, the issue is not limited to whether or not the patient is brain - damaged; it also involves the description of the patient's capabilities and limitations, the kinds of psychological changes being experienced, the impact of these changes on the patient's behavior and personal experience, and their implications for treatment, patient and family counselling, and rehabilitation. The relative degree of deficit, as represented by the individual's actual score, and the distribution of errors throughout a test, are important considerations in the analysis of subtle patterns of test scores for individual patients.

If an abbreviated form of the Category Test is to have practical utility for the individual case, it must be shown that the estimated scores approximate the actual scores with an acceptable degree of accuracy. Of the studies that examined the validity of abbreviated forms, only Sherrill addressed this issue. He found that his C120 prediction equation produced estimated scores that were within a 10-point range of their respective actual scores for 53% of his outpatient sample. Twenty - six percent of subjects had estimated scores which were more than 10 points lower than their actual scores and 21% had estimated scores that were more than 10 points higher than their actual scores. These results indicate only a moderate degree of predictive accuracy when this more stringent validity criterion is applied. Therefore, before either short form could be considered for widespread clinical use, further investigation of the utility of the short form prediction equations is necessary, using the degree to which the estimated scores approximate the actual scores as the validity criterion.

Also, the precision of the predictor in estimating actual scores is influenced by the pattern in which the estimated scores fall around the regression line, as reflected in the error variance. We assume, when adopting the straight line as a model to describe the relationship between the two variables, that the errors have equal variances. However, in a given situation, special circumstances may alter this relationship and the individual scores may deviate around the line of best fit in some specifiable way (Anastasi, 1982, p 158). Therefore, it is important to examine the form, or pattern, of the relationship to determine if, and in what way, deviations from the linear model occur, and whether any such deviations could be used to assist prediction. For example, if the bivariate distribution is heteroscedastic, then predictions may be better at one end of the distribution than the other.

The second condition for justification of an abbreviated form of a test is that it results in a practical saving of time. Of the studies that administered a short form (the C120) to a sample (Sherrill; Gregory et al.), only Gregory et al. reported the time saved. They found that the average administration time for testing normal subjects was cut in half (from 60 - 70 minutes to 30 - 35 minutes) by using the short form. However, experience in our setting over 8 years has been that the Category Test typically takes from 30 to 60 minutes to administer, with the majority of impaired patients completing the test in 40 - 50 minutes. Empirical support for this time estimate comes from a recent study by Finlayson, Sullivan, & Alfano (1987), who also question the administration time of from one to two hours described in the literature. Finlayson et al. found that, despite their generally severe level of impairment, 81% of the patients in two samples took from 20 to 50 minutes to complete the test, and less than 7% took more than 1 hour. Assuming that most patients take 45 minutes to complete the actual test and that use of the C120 would reduce the time by half, the time saved would be approximately 20 minutes. This would be of minor significance in the context of the overall time required to interview the patient and administer the entire Halstead - Reitan Battery and supplimentary measures (approximately 5% if the entire procedure took approximately 6 hours). However, since the Category Test provides immediate feedback as to the individual's success or failure, and the recognition of persistant failure can be very stressful for both patient and examiner, even 20 minutes would represent a useful saving of time in that it would reduce this negative experience for the more impaired patient.

The sample employed in this study consists of adult inpatients referred to the neuropsychology service of a large, tertiary - care general hospital. The sample is neurologically heterogeneous and the majority of patients show diffuse cerebral impairment.

The usual reasons for referral include assistance in diagnosis, monitoring the effects of various treatments, and requests for opinions and recommendations concerning future management and rehabilitation. Requests for neuropsychological assessments to assist in diagnosis usually go beyond the question of whether or not the individual is impaired; subtle distinctions between the kinds of deficits suffered by different diagnostic groups must often be considered. Typical diagnostic issues include: (a) participation in the diagnosis of various diseases that are characterized by some type of cerebral involvement, such as Alzheimer Disease, Pick's Disease, and Multi - Infarct Dementia; (b) identification of secondary cerebral involvement in recognized diseases or conditions (e.g., Parkinson's Disease, Huntington's Disease, Multiple Sclerosis, Systemic Lupus Erythematosis, renal disease, diabetes, hypertension, toxicity, and hypoxia); (c) differential diagnosis (e.g., depressive pseudodementia, pseudoseizures, partial complex seizures); and (d) investigation for evidence of intracranial lesions of various types.

Contributions are also made to the decision - making process with respect to appropriate treatment. For example, concommitant Alzheimer's Disease in a patient with Normal - Pressure Hydrocephalus is a negative prognostic factor in considering treatment with the insertion of a ventriculo - peritoneal shunt. Other common reasons for referral are to monitor the effects of treatment (e.g., medication trials, neurosurgical interventions such as for aneurysm and arterial - venous malformations, and intracranial irradiation), and to monitor recovery over time, as in the case of head injuries, hypoxic conditions, and cerebrovascular accidents.

Provision of comprehensive neuropsychological information to patients and families, as well as to the treatment team, is important in order that they may develop realistic expectations and make appropriate plans. In the clinical situation, neuropsychological test data are used to make inferences about the behavior of individuals in practical terms, taking into account the premorbid history and, as much as possible, the environmental circumstances to which patients are returning. Because of the high sensitivity of the Category Test in identifying difficulties in analyzing problem situations and adapting to circumstances as they unfold, it is one of the most useful instruments in the test battery for making inferences about the individual's behavior in the real world. The complex mental processes that the Category Test appears to measure are often crucial to the person's ability to function effectively in daily living. An individual may have quite severe specific cognitive deficits (e.g., impaired language or visuo - spatial functions) and be able to make a satisfactory adjustment in living so long as these immediate adaptive abilities are relatively intact. On the other hand, a person may demonstrate no obvious specific cognitive deficits and yet be seriously compromised in his or her occupational and interpersonal functioning because of impairment in the type of ability measured by the Category Test.

This investigation has two objectives. The first is to determine which short form provides the best prediction equation for the long form using multiple regression analysis; this provides a constructive replication of Sherrill's (1985) procedure using a more severely impaired inpatient sample. The study also examines the predictive accuracy of the sample prediction equations for the C120 and the C108, and of the equations developed by Sherrill and Calsyn et al., with respect to two validation criteria. The first criterion, the degree of agreement between estimated and actual scores with respect to Halstead's cut - off score, was chosen in order to (a) compare the present results with the results of previous studies, and (b) cross-validate Sherrill's and Calsyn et al.'s prediction equations on a different clinical sample. The study also examined how closely the estimated scores approximated the individuals' actual scores. This type of predictive accuracy has received little or no attention in earlier studies even though important clinical consideration must be given to where the individuals' scores fall on the continuum when making inferences from Category Test scores. In this latter respect, the present investigation makes a unique contribution to the body of studies exploring shortened versions of the Category Test.

METHOD

Subjects:

The files of 225 consecutive inpatients referred for neuropsychological assessment to the Psychology Department of the Victoria General Hospital were reviewed and, of these, the 158 patients who had completed the Category Test were the subjects of the study. For the majority of the remaining 67 patients who did not have complete test protocols, the severity of the patients' impairment had necessitated either prorating the Category Test as per the technique suggested by Reitan (1969), discontinuing it, or replacing it with a less complex measure of concept formation.

Table 1 contains a summary of patient characteristics, referral source and reasons for referral. The subjects' ages ranged from 16 to 72 years (*M* 42.04, *SD* 14.86). They had a mean of 11.3 (*SD* 3.63) years of education (range 3 - 22). Just over 60% were male. These subjects represented a heterogeneous group of referral diagnoses, the majority involving diffuse cerebral impairment.

Table 1

Description of Sample (n = 158)

Sex:	Male	62%
	Female	38%
Age:	*M*	42.0
	SD	14.9
	Range	16-72
Years of Education:	*M*	11.3
	SD	3.7
	Range	3-22
Referral Source:	Psychiatry	53.2%
	Neurology/ Neurosurgery	41.8%
	Other	5.1%
Reason for Referral:	Dementia	18.4%
	Encephalopathy	16.5%
	Depressive Pseudodementia v Dementia	14.6%
	Seizure Disorder	13.9%
	Head Injury	12.0%

Referral Source continued:	Other Psychiatric	10.1%
	Cerebrovascular Accident	5.1%
	Post-Neurosurgical Intervention	5.1%
	Other Neurologic	4.4%

All subjects had been administered the Category Test by the author or one colleague in accordance with the standard procedure described by Reitan (1969). The mean score on the test was 76.66 (*SD* 28.97), with scores ranging from 10 to 134 errors. Twenty - eight subjects (17.7%) had scores which fell below Halstead's cut-off score of 51 errors and 130 subjects, or 82.3%, had scores above 51.

Materials

The Category Test uses a projection apparatus for presentation of 208 slides consisting of combinations of numbers and geometric and other figures. Connected to the apparatus is a set of four levers, numbered from one to four. The subject is instructed to look at the screen, decide which number the picture suggests, and depress the lever having the same number. Depression of the levers causes a bell to sound if the correct answer is chosen or a buzzer to sound if the choice is 'incorrect'. Only one choice is permitted for each slide. The test is divided into seven subtests; the first six each have a single underlying principle throughout, progressing from simple concepts of matching and counting to more complex concepts of uniqueness, position and fractionality. The seventh subtest consists of items from the preceeding subtests and is a measure of recall. The subject's errors are summed for a total score, with 51 errors representing the cut - off score for brain-damage (Reitan, 1966).

Procedure

Each subject's Category Test protocol was rescored by the formulae described by Gregory et al.(1979) and Calsyn et al. (1980), which are listed in Table 2.

Statistical Analyses

Intercorrelations of the variables of age, education, and scores on the standard and two short forms of the Category Test were determined. Using those variables that correlated significantly with the original Category Test as independent variables, and the original Category Test as the dependent variable, multiple regression analyses were undertaken to determine the best regression model for prediction of the original test score. The relationship between the actual scores and the estimated scores generated by the sample equations for the C120 and the C108 were examined through residual analyses. Sherrill's equations for the C120 and the C108, and Calsyn et al.'s equation for the C108 were cross - validated by applying them to the sample data to obtain estimated scores. The relationships between the actual scores and the estimated scores generated by these equations were also examined through residual analysis. The predictive accuracy of each of the five equations was evaluated in terms of the standard error of estimate, Halstead's cut- off score for impairment, and the degree to which the predicted scores approximated the actual Category Test scores.

Table 2

Item Content for Halstead's Category Test (C208) and Proposed Short Forms C120 (Gregory et al., 1979) and C108 (Calsyn et al., 1980)

Subtest in C208	Number of Items in C208 Subtest	Number of Items Retained in C120	C108
I	8	all 8	all 8
II	20	first 16	all 20
III	40	first 32	all 40
IV	40	first 32	all 40
V	40	first 32	0*
VI	40	0	0
VII	20	0	0
Total	208	120	108

* = entire subtest omitted

RESULTS

Determination of the Best Short Form

The intercorrelations among scores for the three forms of the Category Test (standard form and the two short forms) and the subject variables of age and education are presented in Table 3. Both short forms were highly correlated with the standard form and with each other. Of the subject variables, only age was significantly correlated with the test scores and, therefore, retained for further analysis.

Table 3

Intercorrelations Between Three Forms of the Category Test, Age, and Education

	C120	C108	Age	Education
C208	.9529*	.8914*	.4087*	-.1424 *ns*
C120	—	.9419*	.3481*	-.1424 *ns*
C108	—	—	.3150 *ns*	-.1100 *ns*
Age	—	—	—	.0289 *ns*

* $p < .001$, 2 - tailed

Table 4 summarizes the results of the analyses of all possible regression models (see Appendix A for complete results). Of the simple linear models, C120 - based analysis produced an R^2 of .9081, while C108 yielded an R^2 of .7946. When both C120 and C108 were considered, C108 did not account for a significant amount of the variance beyond that explained by C120 (R^2Ch = .0003, F = .582; *ns*). Controlling for C108, C120 accounted for 11.39% (FCh = 192.762, $p < .000$) of the variance, beyond what had already been accounted for by C108. Thus, the C120 was the stronger predictor, accounting for 90.8% of the variance in the standard Category Test, while the C108 on its own accounted for 79.5%. The combination of C120 and age increased the R^2 to .9148, with age

Table 4

R^2 and Standard Error of Estimate (SEE) for All Possible Regression Models

Independent variable included in regression model	R^2	*SEE*
C120	.9081	8.81
C108	.7946	13.17
Age	.4087	26.52
C120, C108	.9084	8.82
C120, Age	.9148	8.51
C108, Age	.9015	12.62
C120, C108, Age	.9151	8.52

explaining an additional 0.67% of the variance (FCh = 12.258; $p < .001$) beyond that accounted for by C120. Inclusion of all three variables resulted in only a negligible increase in R^2 (.9151).

While the contribution of age was significant, it only accounted for an additional 0.67% of the variance and was not considered further. The results of multiple regression analyses supports Sherrill's (1985) conclusion that the C120 is the better of the two abbreviated forms for predicting the long form of the Category Test.

The present sample prediction equations , and those developed by Sherrill (1985) and Calsyn et al. (1980), and their respective standard error of estimates, are shown in Table 5. The *SEE* is the standard deviation of errors of prediction about the regression line. It describes the accuracy with which predictions are made by the regression equation. If it can be presumed that a predicted relationship is linear and homoscedastic, then the *SEE* can be taken as descriptive of the accuracy of prediction at all levels of predictor scores (Ghiselli et al., p 288). In this case, the interpretation of the *SEE* is straightforward and, since the expression of error of estimation is in standard form, it provides a common index for the comparison of results of different studies of a criterion (Wiggins,1973, p 18). However, if there is considerable lack of fit with respect to a linear model, or homoscedasticity cannot be assumed, the *SEE* must be seen as a weighted average of the errors of prediction, and its interpretation becomes more difficult. Sherrill's equations for both short forms had slightly smaller *SEE* 's than the sample equations , suggesting that the comparable short form equations should be of similar accuracy in predicting the actual Category Test score if the assumptions of linearity and homoscedasticity hold. The *SEE* for the Calsyn et al. equation was unavailable for comparison.

Table 5

C120 and C108 Regression Equations and SEE 's for Present Sample and from Previous Studies

Equation	Estimated Score	=	Slope (S.F.)	+	Constant	*SEE*
<u>C120</u>						
Sample	C208'	=	1.532 (C120)	+	8.407	8.81
Sherrill (1985)	C208'	=	1.743 (C120)	+	0.24	7.5
<u>C108</u>						
Sample	C208'	=	1.337 (C108)	+	24.279	13.17
Sherrill (1985)	C208'	=	1.696 (C108)	+	13.93	12.9
Calsyn et al. (1980)	C208'	=	1.4 (C108)	+	15	—

Cross-Validation of C120 and C108 Prediction Equations

The prediction equations developed by Sherrill and Calsyn et al. were applied to the sample. The predicted, or estimated, scores generated by each prediction equation were analyzed using linear model techniques with the predicted scores as the dependent variable and the actual scores as the independent variable. The results of this analysis are given in Table 6. The correlation of Sherrill's equation for the C120 form with the actual form was .9529, and the correlations of both Sherrill's and Calsyn et al's equations for the C108 form were .8914. These correlations were almost identical to those derived in previous studies (See Appendix B).

Table 6

Correlations and SEE 's for Estimated v Actual Scores for Cross - Validated Equations

Equation	*r*	*SEE*
C120 (Sherrill, 1985)	.9529	8.4
C108 (Sherrill, 1985)	.8914	14.89
C108 (Calsyn et al., 1980)	.8914	12.29

The *SEE* for the C120 equation was smaller than the *SEE* 's for the C108 equations, as may be expected since the C120 is the more accurate prediction equation. On cross-validation, Sherrill's C120 equation had a *SEE* of 9.55, just slightly larger than the *SEE* of 8.81 yielded by the sample C120 equation and the *SEE* of 7.5 obtained from Sherrill's original study. This suggests that Sherrill's equation may be almost as good a predictor of the actual Category Test when applied to a second sample, as it was for the original sample.

Among the C108 equations, Calsyn et al.'s equation had a smaller *SEE* than

Sherrill's (12.29 v 14.89). The *SEE* of Calsyn et al.'s equation was also slightly smaller than that of the sample equation for the C108 (i.e., 13.17), suggesting that Calsyn et al's equation may be the best of the C108 prediction equations. The *SEE* of Sherrill's C108 equation, on cross-validation, was similar to that found for his original sample (12.9).

The agreement between estimated scores generated by each equation and the actual scores, with respect to Halstead's cut - off score is given in Table 7.

Table 7

Estimated v Actual C208 Scores with regard to Halstead's Cut - off Score for Impairment

Equation	Scores: Estimated	Actual: ≤ 50	Actual: ≥ 51	Correctly Classified
Sample:				
C120	≤ 50	26	8	86%
	≥ 51	2	122	
C108	≤ 50	26	13	83%
	≥ 51	2	117	
Cross - validated:				
C120 (Sherrill, 1985)	≤ 50	27	11	87%
	≥ 51	1	119	
C108 (Sherrill, 1985)	≤ 50	27	15	84%
	≥ 51	1	115	
C108 (Calsyn et al., 1980)	≤ 50	27	19	81%
	≥ 51	1	111	

The overall correct classification rates were 87% for Sherrill's C120 equation, 84% for Sherrill's C108 equation, and 81% for Calsyn et al.'s C108 equation. False positives and false negatives were evenly distributed for all equations. These classification rates were similar to both those reported in earlier studies and to those produced by the equations of the present sample. The latter "hit rates" were 86% for the C120 equation and 83% for the C108 equation.

These results indicate that the Sherrill and Calsyn et al. equations held up well on cross - validation when the degree of relationship (r) and predictive accuracy in terms of the standard cut - off score are used as the validation criteria.

Examination of Assumptions Relating to Regression Analysis

The above interpretation of validity coefficients and standard error of estimates was based on the assumptions that the relationship between the variables is (1) linear and (2) homoscedastic (i.e., that the variances in the columns of scores are equal to each other and the variances in the rows of scores are also equal to each other). If these assumptions are not met, the correlation coefficient and the standard error of estimate may be misleading with respect to the accuracy of prediction since individual scores may deviate around the regression line in a complex manner. These assumptions may be examined by analyzing the residuals, which are the differences between what is actually observed, and what is predicted by the regression equation. In regression analysis the assumptions relating to errors are that the errors are independent, have zero mean, a constant variance, σ^2, and follow a normal distribution (Draper & Smith,1966, p 86). If these assumptions do not appear to be violated, the residuals can be thought of as the observed errors if the linear model is correct. However, since residuals contain both random and nonrandom components even if the true theoretical model is correctly specified, they can be explored to assess model specifications, model assumptions, and the accuracy of prediction (Gunst & Mason, 1982, p 225).

Model Specification: The Assumption of Linearity : Correct model specification involves two important aspects: that all relevant variables must be contained in the data base, and that the proper functional form of each predictor must be defined in the prediction equation (Gunst & Mason, 1982, p 241). The previous studies of abbreviated forms of the Category Test, and this investigation, are based on the assumption that the simple linear model provides the best explanation of the relationship between scores on the standard Category Test and the respective short forms. The appropriateness of this assumption can be assessed by performing regression analyses of the residuals against the predictor variable (i.e., short form scores) and examining the resultant plots. If the correlation between the residuals and the short form scores equals zero, and there is no discernable trend in the plot, then it can be said that the assumption appears not to be violated and the specification of the model may be correct. The results of these analyses for all of the prediction equations are presented in Table

8.

Table 8

Correlations between Residuals and Predictor Scores and Residual Means for All Regression Equations

Equation	*r*	Mean
Sample:		
C120	0.000	0.000
C108	0.000	0.000
Cross-Validated:		
C120 (Sherrill, 1985)	- .397	- 1.715
C108 (Sherrill, 1985)	- .466	- 4.177
C108 (Calsyn et al., 1980)	- .062	6.399

Both sample equations (C120 and C108) demonstrated zero correlation between the residuals and the predictor scores, and no apparent trend in the respective plots. Similiarly, the correlation of residuals and predictor scores for Calsyn et al.'s C108 equation approached zero (- .062) and there was no discernable pattern in the plot. Thus, the assumption that the simple linear model is an appropriate specification of the relationship between the standard and the abbreviated forms of the Category Test appears not to be violated for the derivation sample. The formula determined by Calsyn et al. also stood up well on cross - validation in this respect.

However, for Sherrill's C120 and C108 equations, the residuals and the predictor scores were correlated at - .397 and .- 466, respectively. This indicates that a linear effect was not removed from the independent variable in each case (i.e., the short forms). Thus, in both cases the fitted model is lacking one (or more) relevant variables needed to properly explain the observed variation in the actual Category Test score. The negative slope of the regression lines indicate that low scores on the short forms may yield negative residuals and high scores positive residuals when these equations are applied in situation outside of those upon which the equations were derived. That is, these equations may show a systematic tendency to produce estimated scores that are higher than the actual scores for those at the lower end of the distribution. For actual scores at the higher end of the continuum, these equations may tend to produce estimated scores that are lower than the actual scores. Inspection of the plots of residuals versus predictor scores for Sherrill's equations revealed no obvious trends.

Another way of examining the specification of the model is to calculate the mean of the residuals (which should always equal zero). The residual means for the two derivation equations were zero, and the residual mean of Sherrill's C120 equation was - 1.70, which was not significantly different from zero. Sherrill's C108 equation yielded a residual mean of - 4.18 and the residuals of Calsyn et al's C108 equation had a mean of 6.40. Violation of the assumption of zero mean for the residuals indicates that the model is not correct, since the residuals contain both random and systematic, or variance error and bias error, components. With respect to the latter two equations, Sherrill's C108 equation shows a systematic tendency to estimate scores higher than the actual scores, while Calsyn et al.'s equation shows a systematic tendency to estimate scores lower than the actual scores, when applied to a second sample.

Homoscedasticity: Another assumption related to regression is that the errors have equal variances, i.e., that they are homoscedastic. When error terms have unequal variances, they are said to be heteroscedastic. If the distribution is heteroscedastic, then predictions may be better at one point in the distribution than at another. Heteroscedasticity can usually be detected by visual examination of plots of the residuals against a variety of explanatory variables. The residuals should appear as points in the form of a horizontal band with no indication of the presence of any systematic trends. The violation of a specific assumption, in this case, homoscedasticity, can sometimes be more evident from one type of plot that from another, so that it is important to plot the residuals against a number of selected variables before passing judgement.

When the squared residuals (e^2) were plotted against the predicted values, heteroscedasticity was evident for each of the equations. Since e^2 reflects the contribution of a given score to the error sum of squares, which is an estimate of σ^2, the plot using e^2 accentuated the types of trends existing between the residuals and the predicted scores for each equation (Gunst & Mason, 1982, p 237).

The plots for both C120 equations revealed a tendency for the error variation to increase as the scores increased, suggesting that these equations may predict low actual scores more accurately than high actual scores. The plot for Calsyn et al.'s equation showed the reverse of this trend, demonstrating decreasing error variation with increasing score. This suggests that, for this equation, prediction may be better for higher actual scores.

The plots for the sample C108 equation and Sherrill's C108 equation demonstrated bimodal distributions. That is, the error variation was larger at both ends than in the middle. Thus, these equations may achieve their most accurate predictions among scores in the middle of the distribution. In addition, the plot for Sherrill's C108 equation indicated that the error variation was largest at the high end of the distribution, indicating that this equation may be least accurate in predicting when the actual score is at the higher end of the distribution.

Implications of Model Misspecification and Heteroscedacticity: The error terms of the three cross validation equations contained both error variance and bias variance components. The presence of the latter indicates that one or more relevant variables may be missing from the model. Therefore, although the linear model may be appropriate for explaining the relationship between the actual and short forms of the Category Test, cross - validation suggests that the univariate form of this model does not fully describe the relationship between the variables.

Although violation of the assumption of homoscedasticity has no effect on the unbiasedness of the estimated regression coefficients, it does affect the precision of these estimates as measured through their variances (Berenson, Levine, & Goldstein, 1983, p 399). A highly important implication of heteroscedasticity is that very few inferences can be drawn from the standard error of estimate, since it now represents simply a summary statement of the distributions of scores in the columns and the rows. Also, the presence of heteroscedasticity means that the factors that determine the relationship between the actual and short forms of the test operate in a complex manner.

Whereas the examination of the residuals from one perspective may suggest the possibility of one type of systematic tendency in the error variation, examination from another angle may indicate the possibility of another, apparently inconsistent trend in the same data. Were the primary goal of the study to examine the statistical relationship between the two variables, these findings would lead one to investigate the addition of other variables and transform aspects of the data to reduce heteroscedasticity and obtain better model specification. However, as the goal of this study was to examine the utility of the short forms of the test in terms of their predictive accuracy, these somewhat ambiguous results indicate that it is not possible to describe the bias components of the prediction equations in a manner that could be used to improve clinical prediction.

Predictive Accuracy with Respect to Individual Scores

Focussing on the specific proportion of correct predictions in a particular situation makes no assumptions about the form of the distributions and takes advantage of the nonlinear as well as linear components of association between two variables (Ghiselli et al., 1982, p 310). Therefore, it is possible that the equations may still produce satisfactory "hit rates" with respect to a specific criterion, despite the violations of assumptions that have been demonstrated. This was the case when Halstead's cut - off score was used as the criterion for predictive accuracy; all equations yielded relatively high "hit rates". In the clinical situation, diagnostic inferences from Category Test scores depend on the consideration of where on the continuum the individual's score falls. Thus, the degree to which the estimated score approximates the actual score on the continuum would be the most clinically relevant criterion for predictive accuracy. Due to the presence of heteroscedasticity, the standard error of estimate cannot be used to estimate the accuracy of individual predictions. Therefore, a high proportion of the predicted scores would have to fall within a relatively narrow range of the actual scores if the equations were to be applicable at the level of the individual.

It was reasoned that variation within 10 points of the actual score would not seriously alter the quantitative inferences that could be drawn from the test score. Therefore, estimated scores falling between plus and minus 5 points of the actual score were considered to be sufficiently accurate to be called well - estimated, or well - predicted.

Subjects were classified as well - estimated if their estimated scores fell within plus or minus 5 points from their actual score. Those whose estimated scores deviated by more than plus or minus 5 points from their actual score were classified as either under - and overestimated. Table 9 shows the proportion of subjects in these three groups for each of the prediction equations.

Table 9

No. and Percent of Estimated Scores which are Underestimates [1], Good Estimates [2], and Overestimates [3] of the Actual Category Test Score

Equation	Estimates					
	Over		Good		Under	
	n	%	*n*	%	*n*	%
Sample:						
C120	43	27	66	42	49	31
C108	55	35	51	32	52	33
Cross-Validated:						
C120 (Sherrill, 1985)	46	29	69	44	43	27
C108 (Sherrill, 1985)	79	50	35	22	44	28
C108 (Calsyn et al.,1980)	32	20	36	23	90	57

1. Estimated score > actual score by more than 5 points.
2. Estimated score within +/- 5 points of actual score.
3. Estimated score < actual score by more than 5 points.

The two prediction equations for the C120 form, and the sample equation for the C108 produced similar numbers of under - and overpredictions, approximately 30% of the sample in each case. Sherrill reported over - and underprediction rates that were only slightly better (26% v 21%) for his C120 equation applied to his original sample. Sherrill's C108 equation yielded almost twice as many scores overpredicted as underpredicted (50% of the sample, v 28%). Calsyn et al.'s equation for the C108 reversed this pattern, resulting in close to three times as many scores being underpredicted as overpredicted (57% v 20%).

If one considers that the goal of prediction, with respect to individual scores, is to minimize the absolute difference between actual and estimated scores, then the direction of the difference is not important and both over - and underpredictions can be thought of as "errors" (Wiggins, 1973, p 61). Sherrill's C120 equation yielded 44% good predictions on cross - validation, as opposed to 53% on the original sample. The sample equation for the C120 had an accuracy rate of 42%. Not surprisingly, the three equations for the C108 short form had much lower accuracy rates, the best being the present sample equation which produced only 32% good predictions. Sherrill's and Calsyn et al.'s C108 equations yielded 22% and 23% "hit rates", respectively.

These results indicate that both short forms, but particularly the C108, appear to have serious limitations in clinical utility. Even when the "best" short form, the C120, is used to predict an individual score, there appears to be less than even odds that the estimated score will approximate the actual score to within a

10-point range.

DISCUSSION

The results of this study illustrate that it is essential to define the validation criteria within the context of the specific purposes for which the predictor is to be used when investigating the criterion - related validity of a measure. The sample prediction equations, and the cross - validated equations, all yielded high validity coefficients and high hit rates with respect to Halstead's cut - off score for impairment, results that were very similar to those of earlier studies. Such positive findings, particularly for the cross - validated equations, could lead the busy clinician, interested in reducing patient stress and costs, to assume that these equations have a broader clinical application if careful consideration is not given to the appropriateness of the validation criteria. The clinical interpretation of the Category Test must go beyond the basic determination of whether or not the individual shows evidence of cerebral dysfunction, as represented by the Halstead cut - off score, to inferences concerning the individual's level of performance. This requires an estimate of the individual's actual score. Therefore, prediction of the individual score would be the main purpose for using a short form in a clinical setting, and the most appropriate criterion for evaluating the short form's validity would be the precision with which its prediction equation estimates the individuals' actual scores.

Both short forms were shown to have serious shortcomings with respect to the precision with which they estimate the actual score of the individual. When good prediction was defined as an estimated score within a 10 point range of the actual score, Sherrill's (1985) CT20 equation had a hit rate of 53% in his original sample, and 44% when applied to a second sample. The present sample equation for the C120 produced only 42% good predictions. The good prediction rates for the C108 equations were 32% for the sample equation, 22% for Sherrill's equation, and 23% for the equation compiled by Calsyn et al. It appears that neither short form meets Anastasi's (1982) first requirement for justification of an abbreviated form of a test, i.e., that it predict the score on the original form with an acceptable degree of accuracy.

This study also reveals some of the difficulties that arise in using regression techniques in actual situations, and the need to examine the way in which the proposed theoretical model describes the relationship between the variables. Regression analysis uses the residuals from the estimated values to estimate the variance of the random error. If the model correctly describes the data, these residuals provide the proper estimate and the standard error of estimate can be used to describe the accuracy of prediction. However, in many practical situations, the correct model is not precisely known, and therefore may be incorrectly specified. When the correct model is not known, it may not be obvious that specification error has occurred (Freund & Minton, 1979, p. 102). The statistical methods that are available for determining "goodness of fit" require an estimate of the "true" error variance, as could be obtained from replicated measurements or, in some instances, from results from previous experiments or data collections. However, in many cases, such as the present one, such information is not available. Until a generally applicable goodness of fit test is developed, goodness of fit must be examined through the more subjective method of residual analysis.

When the prediction equations developed by Sherrill and Calsyn et al. were applied to the data, and the residuals analysed, violation of the assumptions relating to errors suggested that the simple linear model may not provide an adequate description of the relationship between the original and short form scores. The residual means of both Sherrill's and Calsyn et al.'s equations clearly deviated from zero, and correlations between residuals and predictor scores for Sherrill's C120 and C108 equations of - .397 and - .466, respectively, suggested that bias elements were still present in the error variance. In addition, there were indications of the presence of unequal error variance, or heteroscedasticity, for all equations, with the result that the interpretation of the standard error of estimate is limited and, therefore, cannot be used to erect confidence intervals to describe the accuracy of prediction for individual scores.

In conclusion, the use of prediction equations may be justified in testing situations where it is not necessary to predict the specific criterion performance of the individual. However, in the case of the Category Test, neither short form can be recommended for clinical use.

REFERENCES

Anastasi, A. (1982). *Psychological testing.* 5th. ed. New York: Macmillan.

Bak, J. S., & Greene, R. L. (1980). Changes in neuropsychological functioning in an aging population. *Journal of Consulting and Clinical Psychology*, 50, 143-145.

Berenson, M. L., Levine, D. M., & Goldstein, M. (1983). *Intermediate statistical methods and applications: A computer package approach.* Englewood Cliffs, N. J.: Prentice-Hall, Inc.

Calsyn, D. A., O'Leary, M. R., & Chaney, E. F. (1980). Shortening the Category Test. *Journal of Consulting and Clinical Psychology*, 48, 788-789.

Draper, N., & Smith, H. (1966). *Applied regression analysis.* New York: John Wiley & Sons, Inc.

Erickson, R. C., Calsyn, D. A., & Scheupbach, C. S. (1978). Abbreviating the Halstead - Reitan Neuropsychological Test Battery. *Journal of Clinical Psychology*, 34, 922-926.

Finlayson, M. A. J., Johnson, K. A., & Reitan, R. M. (1977). Relationship of level of education to neuropsychological measures in brain-damaged and non-brain-damaged adults. *Journal of Consulting and Clinical Psychology*, 45, 536-542.

Finlayson, M. A. J., Sullivan, J. F., & Alfano, D. P. (1986). Halstead's Category Test: withstanding the test of time. *Journal of Clinical and Experimental*

Neuropsychology, 8, 706-709.

Freund, R. F. & Minton, P. D. (1979). *Regression methods.* New York: Marcel Dekker, Inc.

Fromm - Auch, D. & Yeudall, L. T. (1983). Normative data for the Halstead - Reitan Neuropsychological Tests. *Journal of Clinical and Experimental Neuropsychology*, 5, 221-238.

Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences.* San Francisco: W. H. Freeman & Company.

Golden, C. J. (1976). The identification of brain damage by an abbreviated form of the Halstead - Reitan Neuropsychological Test Battery. *Journal of Clinical Psychology*, 32, 821-826.

Golden, C. J., Kuperman, S. K., Macinnes, W. D., & Moses, J. A. Jr. (1981). Cross- validation of an abbreviated form of the Halstead Category Test. *Journal of Consulting and Clinical Psychology*, 49, 606-607.

Gregory, R. J., Paul, J. J., & Morrison, M. W. (1979). A short form of the Category test for adults. *Journal of Clinical Psychology*, 35, 795-798.

Gunst, R. F., & Mason, R. L. (1980). *Regression analysis and its application.* New York: Marcel Dekker, Inc.

Prigatano, G. P., & Parsons, O. A. (1976). Relationship of age and education to Halstead test performance. *Journal of Consulting and Clinical Psychology*, 44, 527-533.

Reitan, R. M. (1955). An investigation of the validity of Halstead's measures of biological intelligence. *Archives of Neurology and Psychiatry,* 73, 28-35.

Reitan, R. M. (1966). A research program on the psychological effects of brain lesions in human beings. In N. R. Ellis (Ed.), *International review of research in mental retardation.* Vol. 1. New York: Academic Press.

Reitan, R. M. (1967). Problems and prospects in identifying the effects of brain lesions with psychological tests. *Sinai Hospital Journal,* 14, 37-55.

Reitan, R. M. (1969). *Manual for administration of neurological test batteries for adults and children.* Tuscon, Arizona; Author.

Reitan, R. M. (June,1979). Lecture presented at a Reitan advanced workshop in Cincinatti, Ohio.

Sherrill, R. E. Jr., (1985). Comparison of three short forms of the Category Test. *Journal of Clinical and Experimental Neuropsychology,* 7, 231-238.

Taylor, J. M., Goldman, H., Leavitt, J., & Kleimann, K. M. (1984). Limitations of the brief form of the Halstead Category Test. *Journal of Clinical Neuropsychology,* 6, 341-344.

Vega, A. Jr. & Parsons, O. A.. (1967). Cross validation of the Halstead - Reitan tests for brain damage. *Journal of Consulting Psychology,* 31, 619-625.

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality measurement.* Reading, MA: Addison-Wesley.

APPENDIX A

Summary of All Possible Regression Models

Independent Variables	*R*	R^2	*F*	R^2 Ch	*F* Ch	*SEE*
C120	.9529	.9081	1541.419**			8.81
C108	.8914	.7946	603.392**			13.17
Age	.4078	.1670	31.283**			26.52
C120, C108	.9529	.9081	1541.419**	.9081	1541.419**	8.81
	.9531	.9084	768.934**	.0003	.582*ns*	8.52
C108, C120	.8914	.7946	603.392**	.7946	603.392**	13.17
	.9531	.9084	768.934**	.1139	192.762**	8.52
C120, Age	.9529	.9081	1541.419**	.9081	1541.419**	8.81
	.9565	.9148	832.460**	.0067	12.258*	8.51
C108, Age	.8914	.7946	603.392**	.7946	603.392**	13.17
	.9051	.8127	336.355**	.0182	15.034*	12.62
Age, C120	.4087	.1670	31.283**	.1670	31.283**	26.52
	.9565	.9148	832.460**	.7478	1360.928**	8.51
Age, C108	.4087	.1670	31.283**	.1670	31.283**	26.52
	.9051	.8127	336.355**	.6457	534.453**	12.62
C120, C108, Age	.9529	.9081	1541.419**	.9081	1541.419**	8.81
	.9531	.9084	768.934**	.0003	.582*ns*	8.52
	.9566	.9151	553.028**	.0066	12.007**	8.51
C120, Age, C108	.9529	.9081	1541.419**	.9081	1541.419**	8.81
	.9565	.9148	832.460**	.7478	1360.928**	8.51
	.9566	.9151	553.028**	.0002	.418*ns*	8.52

Summary continued

Independent Variables	R	R^2	F	R^2 Ch	F Ch	*SEE*
C108, C120, Age	.8914	.7946	603.392**	.7946	603.392**	13.17
	.9531	.9084	768.934**	.1139	192.762**	8.52
	.9566	.9151	553.028**	.0066	12.007**	8.51
C108, Age, C120	.8914	.7946	603.392**	.7946	603.392**	13.17
	.9051	.8127	336.355**	.0182	15.034*	12.62
	.9566	.9151	553.028**	.1023	185.542**	8.52
Age, C120, C108	.4087	.1670	31.283**	.1670	31.283**	26.52
	.9565	.9148	832.460**	.7478	1360.928**	8.51
	.9566	.9151	553.028**	.0002	.418*ns*	8.52
Age, C108, C120	.4087	.1670	31.283**	.1670	31.283**	26.52
	.9051	.8127	336.355**	.6457	534.453**	12.62
	.9566	.9151	553.028**	.1023	185.542**	8.52

** $p < .000$
* $p < .001$

APPENDIX B

I. *Summary of Derivation and Cross - Validation Studies of the 120 - Item Short Form of the Category Test*

	Derivation	Cross - Validation		
Description of Sample	Gregory et al. (1979)	Sherrill (1985)	Present Study	
Subjects	Normal & 16% Brain-damaged OP's	Mixed Neurologic OP's	Mixed Neurologic IP's	
Number	70	100	158	
Age	18 - 58	15 - 88	*M* 42.04 *SD* 14.86 16 - 72	
Education	Mainly university students	___	*M* 11.31 *SD* 3.63 3 - 22	
Category Test Score	21% > 51	52% > 51 9 - 142	*M* 76.66 *SD* 27.60 10 - 134 82% > 51	
			C120	SC120[1]
Estimated *M* *SD*	___	___	76.65 27.60	77.88 31.40
Correlation (*r*)	.95	.98	.95	.95
Correctly Classified by Cut-off Score	85%	86%	86%	88%
Prediction Hit Rate				
Overestimated	___	26%	27%	29%
Well Estimated	___	53%	42%	44%
Underestimated	___	21%	31%	27%

1. Cross - validation of Sherrill's C120 equation

II. *Summary of Derivation and Cross - Validation Studies of the 108-Item Short Form of the Category Test*

Description of Sample	Derivation Calsyn et al. (1980)	Cross-Validation Calsyn et al. (1980)	Golden et al. (1981)			
Subjects	Alcoholic IP's	Alcoholic IP's	BD & Psych	BD & Psych	N[1] OP's	BD, N, Psych
Number	99	51	38	61	25	60
Age: *M*	45.8		39.3	33.8	30.1	43.68
SD	12.3		13.4	8.9	7.75	13.5
Ed: *M*	12.1		11.4	10.6	12.3	13.9
SD	2.9		3.1	2.9	3.4	3.1
C208 *M*	60.5		87.56	76.55	32.89	77.89
SD	26.9		30.67	29.9	10.87	34.52
C208'[2] *M*	___	___	83.28	69.89	30.27	72.70
SD			28.14	26.46	9.03	28.65
Correlation (*r*)	.89	.89	.88	.87	.83	.91
Correctly Classified by Cut-off Score	86%	84%	87%	87%	100%	92%
Prediction Hit Rate						
Overestimated	___	___	___	___	___	___
Well estimated	___	___	___	___	___	___
Underestimated	___	___	___	___	___	___

1. Normal
2. Estimated score

II. *Summary of Derivation and Cross - Validation Studies of the 108-Item Short Form of the Category Test continued*

Description of Sample		Cross-Validation Studies			
		Taylor et al. (1984)	Present Study		
Makeup		Normals (18%) BD (82%)	Mixed Neurologic IP's		
Number		168	158		
Age:	*M*	44.28	42.04		
	SD	—	14.86		
			16 - 72		
Ed:	*M*	10.82	11.31		
	SD	—	3.63		
			3 - 22		
C208	*M*	70.35	76.66		
	SD		28.97		
			C108[1]	SC108[2]	CC108[3]
C208'	*M*	62.3	76.67	80.34	69.82
	SD	—	25.83	32.74	27.02
Correlation (*r*)		.91	.89	.89	.89
Correctly Classified by Cut - Off Score		87% of BD 66% of N	83%	84%	81%
Prediction Hit Rate					
Overestimated		—	35%	50%	20%
Well Estimated		—	32%	22%	23%
Underestimated		—	33%	28%	57%

1. Equation derived from sample
2. Sherrill's (1985) equation
3. Calsyn et al.'s (1980) equation