The d2 Test of Attention:
Construct validity and extensions in scoring techniques

MARSHA E. BATES[1] AND EDWARD P. LEMAY, JR.[2]
[1]Center of Alcohol Studies, Rutgers University, Piscataway, New Jersey
[2]Psychology Department, Rutgers University, Piscataway, New Jersey

Abstract

The internal consistency and convergent and discriminant validity of the d2 Test, a cancellation test of attention and concentration, was examined in a sample of 364 U.S. adults. Test-taking strategy, new process scores for assessing performance constancy, and relations to gender and education were explored. Results suggested that the d2 Test is an internally consistent and valid measure of visual scanning accuracy and speed. Overall performance scores were related to a proxy measure of test-taking strategy in the expected direction, and new acceleration and deterioration measures exhibited convergent validity. Suggested directions for future research include discrimination of attentional processes that support immediate and sustained visual scanning accuracy and speed, further examination of the impact of test-taking strategies on overall performance measures, and additional construct validity examinations for the new process measures. (*JINS*, 2004, *10*, 392–400.)

Keywords: d2 test, Attention, Learning, Neuropsychological assessment

INTRODUCTION

The construct of attention has been defined in many different ways in the cognitive and neuropsychological literatures. It is currently thought that attention is not a unitary process, but rather that it comprises multiple, dissociable processes dependent in part on the task or situation at hand during measurement, input modalities, stimulus features, behavioral relevance, and the active processes employed to search, shift, focus, and maintain attention (Luck & Vecera, 2002). Accordingly, current literature distinguishes between different types of attention (e.g., selective, focused, and sustained attention), as well as different types of deficits in attention (e.g., neglect, perseveration, distractibility), each testable with unique neuropsychological approaches and associated with distinct attentional models (e.g., Treisman & Gelade, 1980; Posner & Petersen, 1990; Desimone & Duncan, 1995). The most widely used neuropsychological tests of attention (e.g., the Continuous Performance Test, CPT) can distinguish deficits in speed of processing (both mental and sensory), reaction time (motor), and the inter-

action between processing speed and task complexity, but they are frequently inadequate for identifying specific clinical populations (e.g., Halperin et al., 1991). Given that clinical identification of patient populations is central to the neuropsychologist, tests of attention that offer alternative means of classifying attentional deficits may be valuable.

The d2 Test, a cancellation test involving simultaneous presentation of visually similar stimuli, has been proposed as a particularly useful measure of attention and concentration processes (Brickenkamp & Zillmer, 1998). The task is to cancel out all target characters (a "d" with a total of two dashes placed above and/or below), which are interspersed with nontarget characters (a "d" with more or less than two dashes, and "p" characters with any number of dashes), in 14 successive timed trials (Brickenkamp, 1962). The diagnostic utility and construct validity of the d2 Test have been well supported in European samples, yet this test remains relatively unknown in the U.S. (see Brickenkamp & Zillmer, 1998, for a review). The primary aims of the current study were to (1) examine construct validity and internal consistency of the d2 Test using a U.S. adult sample; (2) determine the psychometric characteristics of performance measures that have been proposed; and (3) derive process measures that may be useful for examining performance constancy and test-taking strategies.

Address correspondence and reprint requests to: Marsha E. Bates, Ph.D., Center of Alcohol Studies, Rutgers University, 607 Allison Road, Piscataway, NJ 08854-8001. E-mail: mebates@rci.rutgers.edu

Although the d2 Test was developed prior to contemporary models of attentional processing, it may be a useful alternative to other cancellation tests in common use due to several unique characteristics (see Lezak, 1995; Spreen & Strauss, 1998). For example, according to Desimone and Duncan's (1995) biased competition model of attentional control, target and distracter objects compete for limited processing capacity during visual search, and thus the test taker must selectively attend to relevant stimuli while filtering out irrelevant stimuli in a rapid manner. For a unique target in an array of homogeneous distracters, detection is simple due to a strong competitive bias towards local inhomogeneities (Sagi & Julesz, 1984). The d2 Test, on the other hand, includes nontarget distracters that are visually quite similar to targets (i.e., a "d" with varying spatial configurations of two dashes), thus reducing the competitive advantage of the targets and requiring more complex processing because competition for attention is high.

In addition to bottom-up or stimulus-limited bias, the competition model suggests that top–down control biases competition toward the information that is behaviorally relevant based on the demands of the task (selection of targets over nontargets). Attentional control becomes more computationally challenging when multiple, visually similar objects occur in the visual field because the stimuli will compete for attention (Desimone & Duncan, 1995; Luck & Vecera, 2002). In the d2 Test, due to their physical similarity to the target, distracters would be expected to share in the bias provided by the top–down attentional template. The mental template and neural representation of the target must therefore be complex to differentiate d2 targets and nontargets, and also to allow for detection of varying stimulus configurations of targets (i.e., the target letter "d" with varying spatial configurations of two dashes).

Given the fast-paced repetition of 14 trials without resting, similarities between target and nontarget stimuli, and visual stimulus variations of correct targets, the d2 Test can be used to measure scanning accuracy and speed, as well as learning and test-taking strategies. Its duration and difficulty allow analysis of the participant's ability to achieve, shift, and maintain attention (elements of sustained attention); focus on and select target stimuli (elements of selective attention); improve or worsen with practice; and develop strategic approaches to discriminating between targets and nontargets.

Construct Validity of Scores

To examine the construct validity of preestablished performance scores, intercorrelations among d2 scores were determined, and their relatedness to other measures of neuropsychological functioning was analyzed. Previously examined d2 performance scores (Brickenkamp & Zillmer, 1998) are described in Table 1 and include (1) total number of characters processed (TOT #), (2) total number of errors

Table 1. Abbreviations, descriptions, and calculation of d2 Test measures

In text abbreviations	Standard abbreviations[1]	Description of measures	Computation
TOT #	TN	Total number of characters processed[2]	Sum of number of characters processed before the final cancellation on each trial
O ERR	O	Errors of omission[2]	Sum of number of target symbols not cancelled
C ERR	C	Errors of commission[2]	Sum of number of nontarget symbols cancelled
TOT ERR	E	Total errors[2]	Sum of all errors of omission and commission
% ERR	E %	Percent of errors[2]	Total number of errors divided by the total number of characters processed
TOT CORR	TN-E	Total correctly processed[2,3]	Total characters processed minus total errors made
CONC	CP	Concentration performance[2,3]	Total number of correctly cancelled minus total number incorrectly cancelled
FLUCT	FR	Fluctuation rate (in speed of processing)[2]	Maximum total items processed in a trial minus minimum total items processed in a trial
ERR DIST	ED	Error distribution[2]	Average errors for last 4 trials minus average errors for first 4 trials
ACCEL		Acceleration[4] (increases in speed)	Intraindividual correlation between trial number and speed (Z transformed)
DETER		Deterioration[4] (increases in errors)	Intraindividual correlation between trial number and errors (Z transformed)
STRAT		Strategy index[4]	The total number of characters processed and the percent of omission errors (O ERR/TOT #) were standardized (z scores) and summed. High score = fast with many errors; Low score = slow with few errors

[1]Abbreviations designated in the d2 Manual (Brickenkamp & Zillmer, 1998).
[2]Existing measures examined in convergent and discriminant validity analyses.
[3]Compared for accuracy in determining overall level of performance.
[4]New measures examined in convergent and discriminant validity analyses.

(TOT ERR), (3) errors of omission (O ERR), (4) errors of commission (C ERR), (5) percent errors (% ERR), (6) total number correctly processed (TOT CORR), (7) concentration performance (CONC), (8) fluctuation rate (FLUCT), and (9) error distribution (ERR DIST).*

The overall performance measures of TOT CORR and CONC significantly correlated with several complex measures of attention such as the Symbol Digit Modalities Test (Smith, 1973), which assesses visual scanning, tracking, and sustained attention. Smaller, but statistically significant, correlations were reported for the Stroop Color Word Test, a measure of concentration, distractibility, and response inhibition, and Trail Making Test Parts A and B, which assess complex visual scanning and mental flexibility in addition to attention (Brickenkamp & Zillmer, 1998). These findings support the validity of the d2 overall performance measures as indicators of visual search and attention. Weak correlations with the Wechsler Adult Intelligence Scale–Revised (WAIS–R) Picture Completion and Information suggested that the d2 Test assesses abilities that are relatively independent of those assessed by these performance and verbal subtests of this general intelligence test (Davis & Zillmer, 1998; cited in Brickenkamp & Zillmer, 1998).

Overall Performance Assessment

Selection of an appropriate overall performance measure is crucial, as alternative measures may be differentially influenced by the test-taking strategies implemented by participants. Brickenkamp & Zillmer (1998) reported that two overall d2 performance measures, TOT CORR and CONC, were strongly correlated with TOT # ($r = .95$ and $.72$, respectively) and were more moderately correlated with TOT ERR ($r = -.32$ and $-.65$, respectively) in a large U.S. college sample. The magnitude of the correlations suggests that TOT CORR is highly reflective of processing speed (TOT #) and less reflective of cancellation errors (TOT ERR), whereas CONC represents both errors and speed about equally. The total number of characters correctly processed is a commonly used measure of overall performance, however, the TOT CORR score can be inflated when the respondent skips over many items (see Brickenkamp & Zillmer, 1998 for an illustration). In the current study, the TOT CORR score was compared to a second score, concentration performance, which is presumably not inflated by excessive skipping because it is based on the number of target and nontarget characters cancelled (Brickenkamp & Zillmer, 1998). Each of these overall performance measures was also compared with an index of test-taking strategies (STRAT; see Table 1) to empirically test the notion that CONC is less vulnerable to this source of error than TOT CORR.

---

*The abbreviations used here were constructed to readily convey the meaning of the performance measures to readers and are not the same as the standard d2 Test abbreviations used in Brickenkamp and Zillmer (1998). Both sets of abbreviations are shown in Table 1.

Existing and New Process Scores

Two new process scores, that assess increases in speed and increases in errors across the 14 test trials, were derived to provide information about individual differences in learning, endurance, and performance constancy. A previously proposed measure of systematic changes in accuracy during test administration, error distribution (ERR DIST), is computed by subtracting the average errors for the first four trials from the average errors for the last four trials (Spreen & Strauss, 1998). Given that this score is used to interpret improvement or deterioration over test-taking time, a more sensitive measure would take into account all trials, and would be represented by the degree of linear relationship that TOT ERR has with trial number. Accordingly, scores that represented intraindividual correlations between trial number (1 through 14) and errors, termed deterioration (DETER), and between trial number and speed, termed acceleration (ACCEL), were computed for each participant to assess changes in scanning accuracy and speed over time. Decreases in errors and increases in speed as testing progresses would suggest learning, whereas increases in errors and decreases in speed during test administration could indicate increasing fatigue or an inability to maintain attention.

In summary, the current study examined the internal consistency, and convergent and discriminant validity of d2 performance scores in relation to other cognitive measures, empirically tested whether CONC is a measure of overall performance that is less influenced by test-taking strategies than TOT CORR, and explored two new measures of test-taking process. d2 scores were expected to share most variance with other measures of visual scanning and attention, including the Digit Symbol Substitution Test and the Trail Making Test Parts A and B, but little variance with measures of general intelligence, memory, and abstraction abilities. d2 error scores were expected to be highly correlated with one another and only marginally correlated with TOT #, given that they reflect distinct dimensions of accuracy and speed, respectively. TOT CORR and CONC should be significantly correlated with both total processed (TOT #) and the error scores (TOT ERR, % ERR, O ERR, C ERR) given that these measures involve both speed and accuracy. CONC, however, should have a weaker association than TOT CORR with a proxy measure of skipping as a test-taking strategy. Finally, we explored the validity of new test-taking process measures.

METHOD

Research Participants

The present study used data from one age cohort of the Rutgers Health and Human Development Project (RHHDP; see Bates & Tracy, 1990; White et al., 2001). The initial sample was obtained through a stratified random telephone sampling technique. Eligibility was based upon year of birth and the absence of serious physical or mental handicap,

language barrier, and institutionalization. A total of 1380 adolescents representing three birth cohorts (1961–1963, 1964–1966, and 1967–1969) were initially tested. The sample was predominantly Caucasian (90%), compared to 83% of the New Jersey population at the study's inception (U.S. Bureau of the Census, 1981). The youngest cohort ($N = 364$, 187 women) has been followed longitudinally across five assessment points; data for this study were obtained at the 5th assessment point when the d2 Test was added to the RHHDP neuropsychological test battery. This birth cohort was between the ages of 28 and 32 (4.4% were 28 years old, 33% were 29 years old, 44% were 30 years old, 15.9% were 31 years old, and 2.7% were 32 years old).** See Pandina et al. (1984) for detailed information on sampling.

Measures

d2 test

The standard version of the d2 Test is a one-page paper-and-pencil cancellation test, consisting of 14 rows (trials), each with 47 interspersed "p" and "d" characters (Brickenkamp & Zillmer, 1998). The characters have one to four dashes that are configured individually or in pairs above and/or below each letter. The target symbol is a "d" with two dashes (hence "d2"), regardless of whether the dashes appear both above the "d", both below the "d", or one above and one below the "d". Thus, a "p" with one or two dashes and a "d" with more or less than two dashes are distracters. The participant's task was to cancel out as many target symbols as possible, moving from left to right, with a time limit of 20 s/trial. No pauses are allowed between trials.

The standard equations used to calculate existing processing measures (TOT #, O ERR, C ERR, TOT ERR, % ERR), overall performance (TOT CORR, CONC), and performance variability (ERR DIST, FLUCT) are shown in Table 1 (see also Brickenkamp & Zillmer, 1998). Given the complexity of ACCEL, DETER, and STRAT measures, their computation is described here. The number of errors and the number of target characters processed on each of the 14 trials was calculated. Next, for each participant, the number of errors was correlated with the corresponding trial number (1–14), and the number of characters processed was correlated with the corresponding trial number. These two within-subject correlation coefficients were then z transformed. The z-transformed within-subject correlation of characters processed with trial number was a measure of acceleration (ACCEL), and reflected the degree to which processing speed increased or decreased during test administration. For instance, a score of 0 reflects no systematic change, whereas a negative value reflects a slowing down of processing speed as trials progressed. The z-transformed within-subject correlation of trial errors with trial number

served as a measure of deterioration (DETER), and reflected the degree to which errors increased or decreased throughout test administration. Also see Michela (1990) for a discussion of this technique.

For the strategy index (STRAT), percent of errors and TOT # were standardized in reference to the sample and these two standardized scores were summed. A high score thus reflected relatively faster processing speeds, but accompanied by more missed targets—a skipping strategy, whereas a low score reflected relatively slow processing speeds with few missed targets—a cautious strategy.

Other neuropsychological measures

Convergent and discriminative validity of the d2 scores was examined in relation to performance on the other neuropsychological tests included in the RHHDP neuropsychological battery (Bates & Tracy, 1990). These included a brief version of the Booklet Category Test (BCT, DeFillippis & McCampbell, 1991; Russell & Levy's revision, 1987), the Trail Making Test Parts A and B (Reitan & Wolfson, 1985), the Vocabulary and Verbal Abstraction Tests from the Shipley Institute of Living Scale (Zachary, 1986), the Digit-Symbol Substitution Test, the Block Design Test, and the Digit Span Test (forward and backward) from the Wechsler Adult Intelligence Scale–Revised (Wechsler, 1981), and the Spatial Relations Test from Thurstone's Primary Ability Tests (Thurstone & Thurstone, 1947). The Shipley Abstraction and Vocabulary scores were squared, and Trail Making A and B scores were square rooted, to reduce skew.

Data analysis

Internal consistency of d2 scores was assessed by Cronbach's alpha. Construct validity of existing and new process measures was examined by computing, and testing for significant differences in, the correlations among d2 scores, and by factor-analyzing d2 scores together with test scores from multiple neuropsychological ability domains. The influence of test-taking strategies on measures of overall d2 performance was examined through correlations of the performance measures with STRAT. Finally, ANOVAs were used to test for gender, education, and their interaction effects on d2 performance.

RESULTS

Internal Consistency

The d2 performance subscales exhibited excellent internal consistency (Cronbach's alpha, see Table 2), with the exception of C ERR. It should be noted, however, that participants made few errors of commission, as is typical of cancellation tests (Smith et al., 2002). C ERRs were significantly lower than O ERRs (within participants $t[354] = 19.62$, $p < .0001$), indicating that total error scores (TOT ERR and % ERR) were most reflective of errors of omission ($r = .99$ and .96, respectively).

**Five participants were eliminated from all analyses due to recent (within 24 hr) use of marijuana, phentermine (stimulant for weight loss), methadone, and heavy use of alcohol (12 beers), which could have altered their neuropsychological test performance.

Table 2. d2 Test descriptive statistics

	Mean	*SD*	*Alpha*[1]	*SEM*[1]
TOT #	496.92	75.00	.97	12.99
O ERR	15.93	14.46	.90	4.57
C ERR	1.24	1.78	.61	1.11
TOT ERR	17.18	15.03	.90	4.75
% ERR	.03	.03	.90	.01
TOT CORR	479.74	73.90	.97	12.80
CONC	195.19	37.15	.97	6.43
FLUCT	12.34	4.56	—	—
ERR DIST	−.05	1.28	—	—
DETER	.01	.32	—	—
ACCEL	−.17	.34	—	—

Note. TOT # = Total Number of Items Processed, O ERR = Errors of Omission, C ERR = Errors of Commission, TOT ERR = Total Number of Errors, % ERR = Percent of Errors, TOT CORR = Total Correct, CONC = Concentration Performance, FLUCT = Fluctuation Rate, ERR DIST = Error Distribution, DETER = Deterioration Coefficient, ACCEL = Acceleration Coefficient, SEM = standard error of measurement.
[1]Provided for additive d2 scores only.

Correlations of d2 Subscales

As shown in Table 3, error scores, O ERR, TOT ERR, and % ERR, were highly intercorrelated, as were "speed" scores, TOT #, TOT CORR, and CONC. Error scores were only modestly correlated with the speed scores suggesting that speed and accuracy are relatively distinct dimensions as tapped by the d2 measures. CONC exhibited statistically higher correlations with O ERR, TOT ERR, and % ERR than did TOT CORR (all $p < .01$) suggesting that CONC, as an overall performance measure, is superior to TOT CORR in reflecting both speed and accuracy components of performance. ERR DIST and, to a lesser extent, FLUCT appeared relatively independent of the other d2 scores,

supporting the idea that overall speed and accuracy are distinguishable from consistency in speed and accuracy.

The average score for accuracy deterioration (DETER) was 0, indicating that accuracy was relatively constant across the d2 trials for many participants. The mean acceleration score (ACCEL) was significantly below 0 ($t = −9.19$, $p < .001$), showing that speed of processing slowed across trials. As expected, DETER was significantly correlated with ERR DIST, suggesting convergent validity for this measure of change in accuracy (Table 3). DETER and ACCEL were positively correlated. Those who increased their speed of processing also tended to increase in error rate, reflecting a speed–accuracy tradeoff. DETER and ACCEL were not notably correlated with the other d2 scores, supporting adequate discriminant validity for these two new measures.

Factor Analysis of Cognitive Measures and d2 Scores

The d2 scores and the other measures of cognitive performance were submitted to a principal components factor analysis (eigenvalue > 1 extraction criterion) with an oblique (promax) rotation. Five factors were extracted, conforming to the number suggested by the scree plot, and accounted for 67.83% of the total variance. As shown in Table 4, the first factor, termed Selective Scanning Speed, was defined primarily by TOT #, TOT CORR, CONC, and Digit Symbol. In addition, Trail Making Test Parts A and B had moderate loadings on this factor. The second factor, termed Selective Scanning Accuracy, was comprised of TOT ERR, O ERR, and % ERR. The third factor, termed General Intelligence/Abstraction, consisted of Shipley Institute of Living Abstraction and Vocabulary, Category Test Errors, Block Design, and Trail Making B. The fourth factor, Selective Scanning Deterioration/Acceleration, had high load-

Table 3. Intercorrelations of d2 Test scores

	TOT ERR	% ERR	TOT CORR	O ERR	C ERR	CONC	ERR DIST	FLUCT	DETER	ACCEL
TOT #	.17***	−.01	.98***	.20***	−.12*	.92***	.05	−.23***	−.01	.02
TOT ERR	—	.97***	−.03	.99***	.37***	−.22***	.04	.32***	−.02	−.03
% ERR		—	−.21***	.96***	.42***	−.39***	.02	.39***	−.02	−.04
TOT CORR			—	.00	−.20***	.97***	.04	−.30***	.00	.03
O ERR				—	.26***	−.20***	.03	.30***	−.03	−.02
C ERR					—	−.27***	.05	.19***	.02	−.06
CONC						—	.04	−.36***	.01	.05
ERR DIST							—	−.10	.84***	.37***
FLUCT								—	−.09	−.09
DETER									—	.39***

Note. TOT # = Total Number of Items Processed, TOT ERR = Total Number of Errors, % ERR = Percent of Errors, TOT CORR = Total Correct, O ERR = Errors of Omission, C ERR = Errors of Commission, CONC = Concentration Performance, ERR DIST = Error Distribution, FLUCT = Fluctuation Rate, DETER = Deterioration Coefficient, ACCEL = Acceleration Coefficient.
*$p < .05$.
***$p < .001$.

Table 4. Factor analysis of d2 scores and cognitive measures

	Scanning speed	Scanning accuracy	General intelligence/ abstraction	Scanning deterioration/ acceleration	Immediate memory
TOT#	*.99*	.18	−.09	.01	.09
TOT CORR	*.98*	−.02	−.08	.01	.09
CONC	*.93*	−.21	−.08	.01	.09
DIG SBL	*.75*	−.02	.11	−.01	−.12
TOT ERR	.11	*.99*	−.04	.02	.02
O ERR	.14	*.97*	−.03	.02	.02
%ERR	−.08	*.97*	−.03	.01	.02
SHIPLEY A	.03	−.07	*.75*	.02	−.00
CAT ERR	.06	.03	*−.72*	.00	−.03
BLOCK	−.16	−.00	*.69*	−.01	.27
SHIPLEY V	−.10	−.02	*.64*	.10	.13
TRAIL B	−.32	−.03	*−.63*	−.01	.08
DETER	−.04	−.00	.04	*.92*	−.01
ERR DIST	.00	.06	.01	*.91*	.04
ACCEL	.05	−.02	.03	*.64*	−.05
DSF	.02	.02	.03	.04	*.82*
DSB	.06	−.04	.18	−.04	*.74*
TRAIL A	−.43	−.04	−.47	.01	.36
FLUCT	−.35	.44	.14	−.13	−.11
SPATIAL	.21	.11	.24	−.16	.30
C ERR	−.18	.44	−.03	.04	−.02
% variance	28.35	14.66	10.39	7.94	5.48

Note. Principal components extraction and Promax rotation. Underlined loadings are those above .50. TOT # = Total Number of Items Processed, TOT CORR = Total Correct, CONC = Concentration Performance, TOT ERR = Total Number of Errors, % ERR = Percent of Errors, O ERR = Errors of Omission, C ERR = Errors of Commission, ERR DIST = Error Distribution, FLUCT = Fluctuation Rate, DETER = Deterioration Coefficient, ACCEL = Acceleration Coefficient, DIG SBL = Digit Symbol Substitution Test, TRAIL A = Trail Making Part A, TRAIL B = Trail Making Part B, SPATIAL = Spatial Relations Test, DSF = Digit Span Forward, DSB = Digit Span Backward, BLOCK = Block Design Test, CAT ERR = Booklet Category Test Errors, SHIPLEY A = Shipley Abstraction, SHIPLEY V = Shipley Vocabulary.

ings of ERR DIST, DETER, and ACCEL. The fifth factor, termed Immediate Memory, primarily reflected performance on Digit Span Forward and Backward. These results support the convergent and discriminative validity of the d2 measures, and are consistent with a distinction between overall levels and fluctuations in speed and accuracy.†

Test-Taking Strategy Index and Overall Performance Measures

The proxy measure of test-taking strategies, STRAT, was significantly more strongly correlated with TOT CORR,

---

†A factor analysis involving only the d2 scores replicated the pattern in the factor analysis reported in Table 4, and explained 77.73% of the total variance. O ERR, TOT ERR, and % ERR loaded on an accuracy factor (variance explained: 33.50%). TOT #, TOT CORR, and CONC loaded on a speed factor (25.07%), and ERR DIST, DETER, and ACCEL loaded on an acceleration/deterioration factor (19.16%). Thus, inclusion of the other cognitive performance measures did not alter the internal factor structure of d2 scores. In addition, a factor analysis excluding the new measures (ACCEL and DETER) did not alter the factor structure; ERR DIST loaded by itself on the deterioration factor and the other factors remained nearly identical. The results of these factor analyses are available upon request to the authors.

r (355) = .56 than with CONC, r (355) = .39 ($p < .01$). Also as expected, given the computation of STRAT, it was strongly correlated with both TOT #, r (355) = .72 and O ERR, r (355) = .82.

Gender and Education Effects

d2 Test scores were submitted to 2 (gender: men, women) × 2 (education: low [completed high school or less], high [beyond high school]) factorial analysis of variance (ANOVA). Women and those in the higher education group (see Table 5) performed better on TOT#, TOT CORR, and CONC than did men and those in the low education group. These effects were qualified by significant gender × education interactions. As shown in Table 6, men in the low education group performed more poorly than all others.

Those in the lower education group also had significantly higher % ERR scores, and were significantly different from the higher education group on all the process measures (FLUCT, ERR DIST, ACCEL, and DETER) (see Table 5). The lower education group performed more slowly and made fewer errors as the trials progressed, whereas the

Table 5. Analysis of variance and descriptive statistics for significant gender or education main effects

d2 score	Gender F	Education F	Men	Women	Low education	High education
TOT #	16.04***	15.31***	485.29 (80.05)	508.38 (68.05)	470.82 (78.62)	506.64 (71.32)
TOT CORR	16.44***	19.18***	468.06 (78.31)	491.20 (67.66)	451.19 (75.22)	490.29 (70.65)
CONC	13.66***	19.63***	189.75 (38.32)	200.45 (35.35)	180.60 (35.06)	200.47 (36.53)
% ERR	0.19	5.86*	.04 (.03)	.03 (.03)	.04 (.03)	.03 (.03)
FLUCT	0.37	5.11*	12.42 (4.48)	12.24 (4.65)	13.26 (5.46)	12.00 (4.16)
ERR DIST	1.55	7.83**	−.13 (1.30)	.02 (1.26)	−.38 (1.45)	.06 (1.20)
ACCEL	0.54	6.68*	−.18 (.37)	−.19 (.40)	−.28 (.41)	−.16 (.38)
DETER	0.48	5.36*	−.02 (.34)	.02 (.35)	−.07 (.36)	.02 (.34)
STRAT	5.72*	1.04	−.13 (1.45)	.13 (1.40)	−.14 (1.67)	.06 (1.33)

Note. df between = 1. df within ranged from 346 to 353, due to missing values. *SD* appears in parentheses.
*$p < .05$.
**$p < .01$.
***p < .001.

higher education group showed less deceleration in speed and maintained their initially lower error rate throughout the trials. An examination of percent errors for each trial revealed significant advantages for the higher education group in trials 1 through 3 ($p < .005$), a somewhat smaller effect in trial 4 ($p < .05$), and nonsignificant effects for the remaining trials ($p > .12$). Thus, education differences in error rates were pronounced in early trials and attenuated in later trials. Finally, significant gender differences (see Table 5) were found for STRAT. Women appeared more likely to adopt a skipping test-taking strategy than did men.

DISCUSSION

The d2 Test of attention is a widely used neuropsychological tool in Europe. The present results suggest that the d2 Test is an internally consistent and valid measure of attention in a U.S. sample. The internal consistency coefficients were nearly identical to those previously reported in other countries (Brickenkamp & Zillmer, 1998) and, with the exception of C ERR, occur well within the typical range of .80 to .95 for neuropsychological tests (Mitrushina et al., 1999). The results also supported construct validity for the d2 Test. The factor analysis identified a selective scanning speed factor, which included three primary d2 measures (total number of characters processed, the total number correctly processed, and concentration performance) as well as the Digit Symbol Substitution Test with primary loadings, and FLUCT and Trail Making Parts A and B with moderate loadings. The Digit Symbol Test involves rapid scanning and identification of target items among similar nontarget characters, and thus can be considered a test of focused attention (Spreen & Strauss, 1998). The Trail Making Tests provide information on attention, visual scanning, eye-hand coordination speed, and information processing (Mitrushina et al., 1999). This is similar to the selective attention factor reported by Brickenkamp and Zillmer (1998) and supports the d2's construct validity as a measure of visual search and attention. Although the two measures of overall performance (number correctly processed and concentration performance) were highly correlated with one another and with total characters processed, the new strategy index revealed an important difference between them; number correctly processed was significantly more strongly related to a skipping test-taking strategy than was concentration performance. This result empirically confirms Brickenkamp and Zillmer's (1998) contention that the total number correctly processed is more heavily influenced by a skipping strategy than is the concentration performance measure.

Table 6. Analysis of variance and descriptive statistics for significant education × gender interactions

d2 score	F	Low education		High education	
		Women	Men	Women	Men
TOT #	10.34**	503.80$_a$ (73.59)	441.20$_b$ (71.45)	509.83$_a$ (66.41)	503.00$_a$ (76.65)
TOT CORR	10.50**	483.84$_a$ (73.25)	421.88$_b$ (64.69)	493.53$_a$ (65.91)	486.61$_a$ (75.80)
CONC	8.71**	195.61$_a$ (36.97)	167.12$_b$ (27.19)	201.96$_a$ (34.82)	198.76$_a$ (38.46)

Note. df between = 1. df within ranged from 350 to 353 due to missing values. Means with the same subscript are not significantly different, $p > .05$. *SD* appears in parentheses.
**$p < .01$.

Scanning accuracy, a second component of selective attention (Spreen & Strauss, 1998), was also evident in the scanning accuracy factor that primarily consisted of total errors, errors of omission, and percent errors. An error of omission is a lack of a correct motor response that may reflect a lapse in vigilance, or from a signal detection perspective, the tendency to adopt a strict response criterion for target cancellation. Both of these explanations suggest that errors of omission act as a measure of performance that is distinct from those involving speed of processing, consistent with Brickenkamp and Zillmer's (1998) distinction of drive and control dimensions of performance on the d2 Test. The results of the factor analysis also suggested the discriminant validity of d2 performance measures relative to intelligence, abstraction abilities, and immediate memory. That is, variables loading on the general intelligence/abstraction and memory factors had negligible cross-loadings of d2 accuracy and speed scores.

The new process measures acceleration in processing and deterioration of accuracy across trials, loaded on the scanning deterioration/acceleration factor with error distribution, an existing proxy measure of shifts in accuracy. This finding suggests convergent validity for the process scores, and shows their independence from measures of overall accuracy and speed. It also supports Lezak's (1995) argument that concentration problems may be due to simple attentional disturbance or an inability to *maintain* a purposeful attentional focus, as well as the distinction between focused and sustained attention (e.g., Mateer & Mapou, 1996). The scanning accuracy and speed factors reflected overall attentional disturbances, whereas the deterioration/acceleration factor reflected disruptions in the continued maintenance of attention. The positive correlation between acceleration and deterioration suggests a speed–accuracy tradeoff—those who did not slow their processing speed as trials progressed made more errors, and the continued maintenance of attention appeared to be achieved through reductions in processing speed. However, the moderate correlation between ACCEL and DETER suggests that not all participants maintained attention by reducing speed. Thus, individual differences in negotiating speed and accuracy from one trial to the next were apparent.

Gender and education accounted for a significant proportion of individual differences in performance and strategy in this sample of young U.S. adults. Those with less formal education tended to decrease their processing speed and made fewer errors as trials progressed, although their overall error rate was higher than those with more education. Follow-up analyses revealed that the education advantage in error rates pertained only to the first four of the 14 trials, after which performance for those with less formal education improved to match that of the higher education group. Prior research has also found relationships between d2 Test performance, learning disabilities, and school performance (see Brickenkamp & Zillmer, 1998). In addition, men appeared to utilize a "cautious" test-taking strategy (slow processing with few errors), whereas women appeared to take a skipping strategy. These data suggest that further research is needed to understand how differences in processing speed and strategy may be used as a tool in education-related and clinical assessment, and perhaps intervention.

The results should be considered with respect to study limitations. The sample was moderately large, but was restricted in age range and ethnic composition. Additional research is needed to address whether the results reported here generalize to populations with other demographic characteristics. Further, although the RHHDP neuropsychological battery included tests that were appropriate to examine the convergent and discriminant validity of the d2 Test in relation to multiple cognitive domains, other cancellation type tests of attention such as the CPT were not included. The battery was thus limited in that it could not provide a fine-grained analysis of similarities and differences between the d2 Test and other tests designed to measure the same aspects of attention as the d2. Similarly, the validity of the proposed measures of performance constancy was primarily examined through correlations with existing d2 scores. Including alternative measures of constancy constructs is needed to further establish convergent validity of ACCEL and DETER. The CPT, for example, includes conceptually similar measures (standard deviation of response time and slope of changes in response time) that may serve as useful comparisons to the existing and new d2 measures of change in performance. Finally, although the test-taking strategy measure confirmed predictions regarding differences between CONC and TOT CORR measures, it is most accurately considered as a measure of expected performance outcomes of strategy adoption, rather than a direct assessment of strategy adoption. That is, high scores on the strategy index reflected high processing speed in the context of many missed targets, the expected performance outcome of adopting a skipping strategy. However, this index did not directly assess whether participants chose to adopt a skipping strategy, and it may be the case that the skipping performance pattern is due to factors other than a skipping strategy, such as a tendency to lose one's place during trials. Future studies might experimentally manipulate test instructions so that speed and accuracy vary in their perceived importance. This could encourage adoption of a specific strategy and would provide a more definitive test of the effect of strategies on the overall performance measures.

It is also important to note that we examined a paper-and-pencil test of attention, despite potential advantages of computerized testing, such as increased accuracy of response timing and scoring. At the same time, computerized tests may not be most appropriate for certain populations, such as those with limited computer experience, the elderly, and children. Barkley (1991), for example, reported that pencil-and-paper CPT formats, rather than computer administered versions, have resulted in higher correlations between children's test scores and parent and teacher ratings of attention deficit hyperactivity disorder (ADHD) symptoms. Ultimately, the relative advantages and disadvantages of computer-administered tests must be weighed in both re-

search and clinical settings with respect to the likelihood that the presentation format will mask or reveal important individual differences.

Overall, the present results support the internal consistency, validity, and potential utility of the d2 Test as a component of attention assessment in the US. Although several cancellation tests are currently in use in the U.S., the d2 Test requires substantial attentional processing due to the complexity of the visual stimuli involved. It presents all data simultaneously rather than successively, and is time-limited yet self-paced. These features make the d2 Test a unique and potentially useful neuropsychological tool for identifying and understanding clinical populations that exhibit attentional deficits. The new scores for deterioration, acceleration, and strategy may extend applications of the d2 Test into realms of learning, endurance, and approaches to test taking. The stimulus characteristics of the d2 Test may also be well suited to the basic study of attentional processes. For example, although the predictions of the biased competition model of attention (Desimone & Duncan, 1995) have been examined primarily for successive stimulus displays involving central fixation, a potentially useful extension would be to the case of selective and sustained visual search including eye movements. Future experimental study would allow the attentional processes involved in completing the d2 Test to be understood in the context of current models of visual selection.

ACKNOWLEDGMENTS

REFERENCES

Barkley, R.A. (1991). The ecological validity of laboratory and analogue assessment methods of ADHD symptoms. *Journal of Abnormal Child Psychology*, *19*, 149–178.

Bates, M.E. & Tracy, J.I. (1990). Cognitive functioning in young "social drinkers:" Is there impairment to detect? *Journal of Abnormal Psychology*, *99*, 242–249.

Brickenkamp, R. (1962). *Aufmerksamkeits-Belastungs-Test (Test d2)*. [The d2 Test of attention.] (1st ed.). Göttingen: Hogrefe.

Brickenkamp, R. & Zillmer, E. (1998). *The d2 Test of Attention*. Seattle, Washington: Hogrefe & Huber Publishers.

Davis, K.L. & Zillmer, E.A. (1998). Contrasts between the d2 test of attention and intelligent measures from a normative sample. Paper presented at the 18th Annual Meeting of the National Academy of Neuropsychology, Washington, DC.

DeFillippis, N.A. & McCampbell, E. (1991). *The Booklet Category Test: Research and clinical form manual*. Odessa, Florida: Psychological Assessment Resources.

Desimone, R. & Duncan, J. (1995). Neural mechanisms of selec-

tive visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

Halperin, J.M., Sharma, V., Greenblatt, E., & Schwartz, S.T. (1991). Assessment of the Continuous Performance Test: Reliability and validity in a nonreferred sample. *Psychological Assessment*, *3*, 603–608.

Lezak, M.D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Luck, S.J. & Vecera, S.P. (2002). Attention. In S. Yantis (Ed.), *Steven's handbook of experimental psychology* (3rd ed.): *Sensation and perception* (pp. 235–286). New York: Wiley.

Mateer, C.A. & Mapou, R. (1996). Understanding, evaluating and managing attention disorders following traumatic brain injury. *Journal of Head Trauma Rehabilitation*, *11*, 1–16.

Michela, J.L. (1990). Within-person correlational design and analysis. In C. Hendrick & M.S. Clark (Eds.), *Research methods in personality and social psychology* (pp. 279–311). London, England: Sage.

Mitrushina, M.N., Boone, K.B., & D'Elia, L.F. (1999). *Handbook of normative data for neuropsychological assessment*. New York: Oxford.

Pandina, R.J., Labouvie, E.W., & White, H.R. (1984). Potential contributions of the life span developmental approach to the study of adolescent alcohol and drug use: The Rutgers health and human development project, a working model. *Journal of Drug Issues*, *14*, 253–268.

Posner, M.I. & Petersen, S.E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

Reitan, R. & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological battery: Theory and clinical implications*. Tucson, Arizona: Neuropsychology Press.

Russell, E.W. & Levy, M. (1987). Revision of the Halstead Category Test. *Journal of Consulting and Clinical Psychology*, *55*, 898–901.

Sagi, D. & Julesz, B. (1984). Detection versus discrimination of visual orientation. *Perception*, *13*, 619–628.

Smith, A. (1973). *Symbol Digit Modalities Test*. Los Angeles, California: Western Psychological Services.

Smith, K.J., Valentino, D.A., & Arruda, J.E. (2002). Measures of variations in performance during a sustained attention task. *Journal of Clinical and Experimental Neuropsychology*, *24*, 828–839.

Spreen, O. & Strauss, E. (1998). *A compendium of neuropsychological tests* (2nd ed.). New York: Oxford University Press.

Thurstone, L.L. & Thurstone, T.G. (1947). *SRA Primary Abilities*. Chicago, Illinois: Science Research Associates.

Treisman, A.M. & Gelade, G. (1980) A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

U.S. Bureau of the Census (1981). *Current population survey: Money, income, and poverty status of families and persons in the United States: 1980*. Current Population Reports, Series P-60, No. 127.

Wechsler, D. (1981). *Wechsler adult intelligence scale–revised*. New York: Harcourt Brace Jovanovich.

White, H.R., Bates, M.E., & Buyske, S. (2001). Adolescence-limited versus persistent delinquency: Extending Moffitt's hypothesis into adulthood. *Journal of Abnormal Psychology*, *110*, 600–609.

Zachary, R.A. (1986). *Shipley Institute of Living Scale: Revised Manual*. Los Angeles, California: Western Psychological Services.