Journal of Clinical and Experimental Neuropsychology
2000, Vol. 22, No. 6, pp. 720-730

1380-3395/00/2206-720$15.00

A Personal Floor Effect Strategy to Evaluate the Validity of Performance on Memory Tests*

Richard I. Frederick
U.S. Medical Center for Federal Prisoners, Springfield, Missouri

ABSTRACT

Four methods of assessing the validity of performance on a word recognition test were compared among 609 criminal defendants engaged in competency-to-stand-trial evaluations. One of the methods, the ''normative'' floor effect strategy, involves comparing an individual's performance to the average performance of individuals with true memory impairment. In this sample, 16.9% of defendants performed below the normative floor for individuals with true impairment. Another method, the ''personal'' floor effect strategy, identifies performance as suspect when individuals perform below a level for which they themselves have already demonstrated intact ability. In this sample, the personal floor effect strategy identified fewer instances of suspicious performance (15.6%), but the strategy may be less sensitive to true memory impairment than the normative floor effect. Consequently, the personal floor effect strategy may be more compelling as evidence of poor effort or bad intentions on memory testing. Convergent validity of the personal floor effect strategy is demonstrated over four analyses.

Criminal defendants must demonstrate intact cognitive capacities in order to resolve the case against them, either by pleading or by standing trial. The U.S. Supreme Court held that all defendants must possess a factual, as well as rational, appreciation of their circumstances in order to be competent to stand trial (Dusky, 1960). The standard for competency in the federal courts is that defendants must be aware of the nature and potential consequences of the charges against them and must be able to properly assist their attorneys in their own defense.[1] These standards apply in criminal courts throughout the United States (Melton, Petrila, Poythress, & Slobogin, 1997).

Forensic psychologists often administer tests of general cognitive capacity and conduct direct examinations of memory capacity (Borum & Grisso, 1995; Heilbrun, 1992) as relevant measures of a defendant's capacities. Barring the influence of certain delusional states, evidence that these basic cognitive capacities are intact is typically predictive of an ability to assist their attorneys and to appreciate their circumstances. Evidence of impaired memory ability, however, may lead to the conclusion that the defendant cannot provide the assistance the attorney requires to evaluate the completeness or accuracy of witness statements, cannot track decisions throughout the process, or cannot maintain alertness and awareness at critical moments. A finding of incompetency may result in an outcome which is favorable to a defendant, including a postponement of trial (which may lead to a

[1] 18 U.S. Code Section 4241(a).

* The conclusions and opinions reported in this paper do not necessarily represent the opinions of the Federal Bureau of Prisons or the U.S. Department of Justice.

Address correspondence to: Richard I. Frederick, Department of Psychology, U.S. Medical Center for Federal Prisoners, Springfield, Missouri, USA. Tel.: ++ 1 417 862 7041 x174. Fax: ++ 1 417 865 4161. E-mail: rfrederi@ipa.net.

Accepted for publication: February 3, 2000.

lower likelihood of successful prosecution), the introduction of tenable mental state defense evidence, or even a decision not to prosecute.

Some individuals who undergo evaluation attempt to influence the conclusions of the evaluator and gain any or all of these potential benefits by choosing to demonstrate only a limited extent of their memory capacity, by actively pretending to be unable to remember new information, or by making false claims of amnesia. In recognition of these factors, researchers have developed methods to evaluate the validity of complaints about memory functioning. For example, Frederick, Carter, and Powel (1995) proposed a method to evaluate suspicious claims of amnesia directly (see also Denney, 1996; and Frederick & Denney, 1998).

Evaluation of the validity of general memory functioning has a much longer history. Rey (1941) proposed a ''performance curve strategy'' (Rogers, Harrell, & Liff, 1993), comparing performance on comparatively easier word recognition memory tasks with more difficult word recall memory tasks. Because recognition memory is typically much stronger than recall memory for most individuals (e.g., see Robinson & Johnson, 1996), individuals who are exerting their best effort should easily recognize more words than they are able to recall without prompting. Rates of misrecognition (i.e., endorsement of a foil) should be far below rates of accurate recognition. Conversely, individuals pretending to be impaired might inadvertently recall more words than they recognize on memory tests. Although this approach has intuitive appeal, Frederick, Sarfaty, Johnston, and Powel (1994) found that the technique was relatively insensitive, identifying less than 5% of malingerers in an analog study. Greiffenstein, Baker, and Gola (1996) reported a sensitivity of 58% in a clinical sample, when the question was whether patients recalled as many or more words than they recognized. This comparison proved to be the least sensitive approach of a variety of measures they examined.

Hiscock and Hiscock (1989) reported a procedure to evaluate memory performances using a two-alternative forced-choice (2AFC) test. In the Hiscock procedure, test-takers memorize a 5-digit number. The examiner presents two 5-digit choices on a recognition task (the target and foil). The examinee must choose which number had been presented for memorization. On tests like the Hiscock procedure (e.g., the Portland Digit Recognition Test; Binder, 1990), the original detection strategy was to compare an examinee's performance to that expected by chance (i.e., random responding). Performances significantly worse than chance provided evidence of feigned or exaggerated impairment, and performances consistent with chance were considered inconclusive. This strategy proved to be relatively insensitive; consequently, although forced-choice testing of memory abilities has become prevalent, the primary strategy of detection for most of these tests has shifted from a comparison versus chance pattern to a *floor effect strategy* (Binder, 1993; Tombaugh, 1997; Slick, Hopp, Strauss, Hunter, & Pinch, 1994).

Floor effect detection strategies (Rogers et al., 1993) compare a score to the average score of individuals who are demonstrably impaired (which establishes the floor). Performance beneath the floor is deemed suspicious. The floor in some 2AFC tests is well ''above'' the range of random response. For example, on the Test of Memory Malingering (TOMM; Tombaugh, 1997), performance is considered predictive of malingering when fewer than 90% of the trials have been correctly answered.

Frederick, Crosby, and Wynkoop (2000) combined a floor effect strategy with a performance curve strategy. They compared performance on a cognitive task to an individual's *own* floor. That is, the highest demonstrated capacity of the individual to consistently and correctly respond to test items was established as the individual's ''floor.'' Failure of the test-taker to respond perfectly or near-perfectly for test items below their own floor strongly supported a conclusion of inadequate effort. In this paper, this method will be referred to as a *personal floor strategy*, as opposed to a *normative floor strategy* (establishing the floor based on the average performance of a group of impaired individuals). Rey's performance curve detection method can also be construed as a personal floor strategy. The floor for the recognition memory task

is established by the individual's own performance on the recall task. Performance on the comparatively easier recognition memory task should be at that level or higher.[2] Greiffenstein et al. (1996) reported that this strategy demonstrated 62% sensitivity and 86% specificity in a clinical sample. The primary goal of the present paper was to compare the normative floor effect strategy and the personal floor strategy in identifying invalid performances on memory testing in a criminal forensic sample.

METHOD

Participants

Participants were 609 male criminal defendants who had been ordered by the federal district courts to undergo evaluation regarding their competency to stand trial (n = 497) or who had been declared incompetent to stand trial and committed for treatment to restore competency (n = 112). Evaluation of *current* cognitive capacity in an adversarial setting with potential punitive outcomes inherently includes a motivation to feign cognitive impairment. Some other evaluations concerning these defendants were completed concurrently, including criminal responsibility evaluations (n = 288), risk assessments (n = 8), and examinations of need for mental health treatment as part of sentencing (n = 16). Mean age was 36.2 (range = 18 to 72 years, SD = 11.0) and mean years of education was 10.8 (range = 0 to 20 years, SD = 3.1). Defendants included 358 White (58.8%), 176 Black (28.9%), 42 Hispanic (6.9%), 22 Native American (3.6%), 5 Asian (0.8%), 2 Pacific Islander (0.4%), and 4 individuals of another ethnicity (0.8%). All defendants were undergoing evaluation on mental health wards in a correctional setting.

Instruments

Instruments included two memory tests, a word recall test to establish a personal floor and a word recognition test to assess the validity of performance.

Auditory Verbal Learning Test (AVLT; Lezak, 1995; Rey, 1941)

The number of correctly recalled words on the first trial of the AVLT established each defendant's personal floor for recognition capacity.

Word Recognition Test (WRT; Lezak, 1995; Rey, 1941)

The WRT consists of 15 stimulus words which are read aloud to the participant. Rey's recognition task consists of visually presenting 30 words (15 targets and 15 foils) among which the examinee must circle words that were presented for memorization. Frederick et al. (1994) adapted the WRT to a 2AFC task, orally presenting each word on the recognition task and requiring a decision at each word (i.e., ''yes'' or ''no'') with respect to whether the word was on the original memory list. Examinees are required to give a response, by guessing if necessary. In this type of administration, the *WRT total score* can be derived by subtracting the number of incorrect recognitions from the number of correct recognitions. This generates a range of possible scores from –15 to 15. The expected mean score for random responding is 0. At 95% confidence, the range of random responding includes scores from –5 to 5.[3] Scores beyond this range, in either direction, reflect the presence of memory ability. Greiffenstein et al. (1996) reported that a WRT total score of 5 or lower is unusual for individuals who are considered to have real memory impairment (normative floor effect).

Three tests that assess feigned neuropsychological impairment were used as independent variables to categorize performance.

Validity Indicator Profile (VIP; Frederick, 1997)

The VIP is a measure of response validity intended to be administered concurrently within a battery of cognitive tests and consists of two subtests, nonverbal and verbal. Based on performance curve characteristics (Frederick, 1997; Frederick, Crosby, & Wynkoop, 2000), performance on the VIP is classified as ''malingering,'' ''irrelevant,'' ''careless,'' or ''compliant.'' Performance classified as

[2] For this paper, a performance curve strategy refers to comparing performance on easy and hard tasks. A floor effect strategy examines failure to meet a minimal performance expectation.

[3] This range is determined by solving the equation: $z = \frac{(x+/-0.5)-Np}{\sqrt{Npq}}$, where z = +/– 1.96 (the magnitude of z at alpha = .05), x represents the score for which z is computed, Np represents the expected mean (in this scoring scheme, Np = 0), and $\sqrt{Npq}$ represents the expected standard deviation of the distribution of scores. N = 30 trials. $p = q = 0.5$. Add 0.5 to x if x < Np; subtract 0.5 from x if x > Np. See Siegel (1956).

malingering strongly indicates a sustained effort to respond incorrectly. Irrelevant performance indicates that performance was independent of average item difficulty. In the absence of bona fide severe impairment, this is considered evidence of a token effort to respond incorrectly. Careless performance indicates an intention to answer items correctly, but effort has likely been compromised or diminished as compared to compliant responding. When performance is classified as compliant, there is evidence of a sustained effort to answer items correctly. Only categorizations by the nonverbal subtest were considered for these analyses; only a relatively small proportion of defendants had been evaluated with the current form of the verbal subtest.

Rey 15-Item Memory Test (RMT; Lezak, 1995; Rey, 1958)
The RMT consists of a card with five rows of three items that appear in a familiar logical sequence (e.g., ''1, 2, 3'' or circle, square, triangle). Defendants were told to remember all 15 items presented to them during a 10-second exposure. After the stimulus items were removed, and following a 10-second delay, defendants were told to reproduce the items in the correct order on a blank sheet of paper. The score consistently reported in the literature is the number of correctly recalled items (Frederick et al., 1994; Guilmette, Hart, Giuliano, & Leininger, 1994). Scores range from 0 to 15; lower scores are consistent with an intention to perform poorly. One defendant did not complete the RMT.

Dot Counting Test (DCT; Rey, 1941)
The DCT consists of twelve 3" x 5" cards on which have been placed either random (ungrouped) or patterned (grouped) dots. The cards were presented to examinees who were instructed to count the dots as quickly as possible without making mistakes. Binks, Gouvier, and Waters (1997) reported the most discriminating score was the number of cards for which an incorrect sum was reported. The number of occasions of miscounts range from 0 to 12; higher scores indicate a motivation to perform poorly. Six defendants did not complete the DCT.

Procedure
Generally, the order of presentation was to first administer the VIP in a group setting. Individual testing followed, beginning with the AVLT. The WRT was administered after completion of other tests, including the DCT and RMT. This order of administration of the two memory tests proved optimal as it precluded the intrusion of WRT words in the AVLT recall trials.

ANALYSES

Four analyses examine the utility of the personal floor effect in evaluating the effort and motivation of defendants who have motivation to feign cognitive impairment. The first analysis compares different strategies for evaluating performance on the WRT. The second analysis introduces a ''memory comparison score'' (MCS), a measure of whether defendants have exceeded their personal floor on the WRT. In the third and fourth analyses, the effect of differences in motivation and effort, as measured by the MCS and independent of true memory capacity, are examined for the RMT and DCT.

Analysis 1. Detection Strategy Effectiveness

Four methods of evaluating the memory performance of defendants were considered. The first method, *below chance responding*, resulted in a positive finding when the WRT score (recognitions minus misrecognitions) was less than –5. The second method, Rey's original *performance curve strategy*, resulted in a positive finding when the number of words recalled on the AVLT exceeded the number of words recognized on the WRT. Positive findings for the third method, the *normative floor effect*, were based on WRT scores of 5 or lower. The fourth method, the *personal floor effect* strategy, resulted in a positive finding when the number of words recalled on AVLT trial one exceeded the WRT total score.

Results
Table 1 shows the frequency of positive scores for each method. Fewer than 1% of defendants scored below chance on their WRT total. The performance curve strategy identified 6.4% of individuals as invalid responders. The normative floor effect and the personal floor effect identified 16.9% and 15.6% of performances, respectively, as questionably valid.

Table 1. Comparison of Strategies to Identify Invalid Performance on the Word Recognition Test (WRT).

Detection Method	Number of Positive Findings	
Below Chance Responding	2	(0.3%)
Rey's Performance Curve	39	(6.4%)
Personal floor	95	(15.6%)
Normative floor	103	(16.9%)

Note. $N = 609$ defendants. Positive findings were defined as follows: *Below Chance Responding*: WRT score < –5 (WRT score is computed by subtracting total misrecognitions from total recognitions); *Rey's Performance Curve*: WRT recognitions < Auditory Verbal Learning Test trial 1 recall (AVLT); *Personal floor*: WRT score < AVLT; *Normative floor*: WRT score < 5.

Comparing Personal and Normative Floor Effect Strategies

Table 2 reflects the high rate of agreement between these two floor effect strategies (kappa = .78, $p < .05$). Fourteen individuals received a positive score on the personal floor effect strategy, but not on the normative floor effect. Twenty-two individuals received positive scores by the normative floor method, but not by the personal floor method. The mean number of words recalled on the first trial of the AVLT by individuals who were identified as probably invalid by the normative floor effect only was 2.6 ($SD = 1.0$, $n = 22$). AVLT performance by individuals identified as invalid by the personal floor effect only was far superior ($M = 7.3$, $SD = 0.9$, $n = 14$; $t = –14.2$, $df = 34$, $p < .05$, Cohen's $d = 5.0$). Consequently, the influence of genuinely poor memory skills cannot be ruled out for individuals identified as questionably valid by the normative floor effect. The normative floor effect strategy may potentially be overly sensitive, incorrectly categorizing impaired individuals as probable malingerers (Guilmette et al., 1994; Lee, Loring, & Martin, 1993; Schretlen, Brandt, Krafft, & Van Gorp, 1991). That is, normative floor effect strategies run the risk of false positive misclassification because they are often established by balancing sensitivity and specificity, instead of emphasizing specificity over sensitivity (Rogers, Bagby, & Dickens, 1992).

When performance is below a floor established by a normative group, severe impairment, bad intentions, or poor effort serve as competing explanations for why the performance is below the floor. But when individuals have themselves demonstrated a capacity to respond correctly at a certain level, it is difficult to explain how performing below that level on an easier task could be construed as evidence of good intentions and strong effort. Poor performance below a personal floor cannot be properly construed as a result of severe impairment. The advantage of the personal floor effect is that the floor is established by the individual taking the test and not by a normative group. The following analyses

Table 2. Cross-tabulation of Positive and Negative Findings on Two Floor Effect Strategies.

Normative Floor	Personal Floor	
	Negative	Positive
Negative	492	14
Positive	22	81

Note. $N = 609$ defendants. Kappa = .78. Positive findings were defined as follows: Personal floor: Word Recognition Test (WRT) total score < Auditory Verbal Learning Test trial 1 recall. Normative floor: WRT total score < 5. WRT total score = WRT recognitions minus incorrect recognitions.

consider the utility of the personal floor effect as an indicator of cooperativeness.

Analysis 2. Comparing Personal Floor Effect with other Validity Measures

A *memory comparison* score (MCS) was computed to evaluate the personal floor effect:

> MCS = [(WRT recognitions) – (WRT misrecognitions)] – (AVLT trial 1 recalls)

This scoring scheme results in positive scores (MCS_{pos}) when recognition performance exceeds recall performance, suggesting motivation to perform well and good effort. A large number of misrecognitions or a small number of recognitions in the presence of a larger number of recalled words results in a negative score (MCS_{neg}), suggesting a motivation to perform poorly or poor effort.

Defendants were categorized by presumed increasing levels of cooperation and effort based on their performances on the RMT, DCT, and VIP. The VIP categorized performance into four categories: Malingering, irrelevant, careless, and compliant. RMT performance was divided into five categories, based on the total score for each test (0-5 items, 6-8 items, 9-11 items, 12-14 items, and 15 items). These categories roughly correspond to ''fewer than two rows,'' ''fewer than three rows,'' ''fewer than four rows,'' ''fewer than five rows,'' and ''five rows.'' DCT performance was also divided into five categories (miscounting 9-12 cards, 6-8 cards, 3-5 cards, 1-2 cards, and 0 cards). For each test, two categories represented a high likelihood of an intention to perform poorly. For the RMT, these categories were for reproducing only 0-5 items or 6-8 items. For the DCT, these categories reflected counting errors on at least half the trials, miscounting 9-12 cards or 6-8 cards. Mean MCS was computed for each category (Table 3).

As performance improved on the RMT and DCT, mean MCS improved as well (Table 3; $r_{RMT} = .37$, $n = 608$, $p < .05$; $r_{DCT} = .35$; $n = 603$, $p < .05$). Negative and zero-order mean MCSs were observed for the categories of RMT and DCT most likely corresponding to a high likelihood of intent to perform poorly. Comparisons among VIP categorizations revealed similar findings. Individuals who exerted strong efforts to fail test items on the VIP (malingering)

Table 3. Mean Memory Comparison Scores (MCS) by Category of Performance on Other Measures of Validity.

Test Category	*n*	*M*	(*SD*)
Dot Counting Test			
Number miscounts			
9-12	17	–2.8	(3.3)
6-8	46	0.6	(4.3)
3-5	195	3.0	(3.6)
1-2	253	3.7	(3.1)
0	98	4.3	(2.8)
Rey Memory Test			
Number items reproduced			
0-5	11	0.0	(3.5)
6-8	52	0.6	(4.3)
9-11	108	2.1	(3.7)
12-14	195	3.4	(3.3)
15	242	4.4	(2.9)
Validity Indicator Profile			
Category assigned			
Malingering	8	–2.1	(4.8)
Irrelevant	74	0.7	(4.3)
Careless	179	2.6	(3.4)
Compliant	348	4.1	(3.0)

Note. MCS is computed as: Word Recognition Test total score – Auditory Verbal Learning Test first trial recall.

earned large magnitude negative MCS (rank-order correlation = .32, n = 609, $p < .05$). Individuals who apparently did not intend to respond correctly on the VIP, but exerted less effort to fail items (irrelevant responding) earned low magnitude positive MCS. Individuals who were motivated to respond correctly on the VIP (careless responders), but may not have exerted full effort earned larger magnitude positive MCS. Finally, individuals who demonstrated strong effort to respond correctly to VIP items (compliant responders) earned the largest magnitude positive MCS.

Analysis 3. Comparisons when Controlling for Memory Capacity

These findings support the assumption that individuals who are feigning impairment or who are motivated to perform poorly earn negative or low magnitude positive MCS. The limitation to this conclusion is that categorizations on the RMT, DCT, or VIP may have been influenced by real memory capacity differences among individuals which, in turn, produced the observed differences in mean MCS. This limitation was addressed restricting comparisons of MCS to individuals who had been matched on demonstrated recall capacity. Individuals were matched if they had equal scores on the first recall trial of the AVLT and if their WRT scores deviated equally from the AVLT score *in different directions.*

For example, two individuals were potentially selected for comparison if they both recalled 5 items on the AVLT. If one individual earned a WRT score of 7 (2 points *above* the recall score), and if the other individual earned a WRT score of 3 (2 points *below* the recall score), then the two individuals were selected for comparisons. Matched individuals (n = 170) were placed in comparison groups according to the negative or positive valence of MCS (MCS_{neg}, n = 85, M = –2.9, SD = 2.0, range = –9 to –1; MCS_{pos}, n = 85, M = 2.9, SD = 2.0, range = 1 to 9). Forming groups in this manner allowed comparisons between DCT and RMT with full assurance that differences in performance were not related to differences in memory ability.

MCS_{neg} individuals were hypothesized to have shown poor effort on memory tests or intended to perform poorly. MCS_{pos} individuals were assumed to have intended to perform well, with higher memory comparison scores reflecting greater effort. It is probable that MCS_{neg} individuals could have produced higher AVLT scores than they actually did. This presented no limitation to the comparison, given that the goal was to ensure only that MCS_{neg} individuals had at least as much memory capacity as MCS_{pos} individuals.

Results

The mean AVLT score for each group was 4.6 (SD = 1.6). Table 4 demonstrates that MCS_{neg} individuals (M = 9.9, SD = 3.0) reproduced significantly fewer items on the RMT than MCS_{pos} individuals (M = 11.9, SD = 3.2; t = –4.3, df = 168, $p < .05$, Cohen's d = .66). MCS_{neg} individuals made more counting errors (M = 4.4, SD = 3.0) than MCS_{pos} individuals (M = 2.7, SD = 2.0; t = 4.2, df = 166, $p < .05$, Cohen's d = .67). These differences cannot be explained on the basis of any memory capacity differences between the two groups. Consequently, an appropriate interpretation for a negative MCS is that it most likely represents an instance of invalid performance either through poor effort or through an intention to perform poorly.

Analysis 4. Interpretation of Magnitude and Valence of MCS

The final analysis considered the meaning of magnitude and valence in interpreting MCS. Memory comparison scores were categorized based on ranges. Scores in the range of –4 to –1 were labeled as ''weak evidence of motivation to perform poorly.'' Scores of –5 and below were labeled as ''strong evidence of invalid performance.'' Scores of 0 to 3 were labeled as ''weak evidence of valid performance,'' and scores 4 and higher were labeled as ''strong evidence of valid performance.'' At each AVLT score (e.g., AVLT = 1), four individuals with different MCS scores, one for each category of MCS, were selected until any one category could not be represented by the remaining individuals with that AVLT score. A total of 68 de-

Table 4. Comparisons of Positive and Negative Memory Comparison Scores (MCS) When Controlling for Memory Capacity.

Test	Positive MCS			Negative MCS		
	n	*M*	(*SD*)	*n*	*M*	(*SD*)
Rey Memory Test (RMT)	85	9.9	(3.0)	85	11.9	(3.2)
Dot Counting Test (DCT)	84	4.4	(3.0)	84	2.7	(2.0)
Auditory Verbal Learning Test	85	4.6	(1.6)	85	4.6	(1.6)

Note. Comparisons for the RMT were significant: $t = -4.3$, $df = 168$, $p < .05$, Cohen's $d = .66$. Comparisons for the DCT were significant: $t = 4.2$, $df = 166$, $p < .05$, Cohen's $d = .67$

fendants were selected; 17 represented each MCS category. Performances on the RMT and DCT were evaluated by category.

Results

Mean AVLT performance for each memory comparison score category was 4.4 words recalled (range = 2 to 6, $SD = 1.4$). Table 5 shows the mean score values for MCS, RMT, and DCT. Significant differences existed among RMT scores, F (3, 64) = 10.7, $p < .05$, and among DCT scores, F (3, 64) = 13.0, $p < .05$. Post hoc comparisons (using Fisher's LSD method, overall alpha = .05) for both the RMT and DCT revealed that the categories of ''strong evidence of invalid performance'' and ''weak evidence of invalid performance'' were significantly different from each other and from any other category. The categories of ''weak evidence of valid performance'' and ''strong evidence of valid performance'' were not significantly different from each other.

The lack of meaningful differences in the ''strong'' and ''weak'' categories for evidence of valid memory test performance may reflect a lack of experimental power. When all individuals with MCS > 0 are examined, however, the correlation between MCS and items reproduced on the RMT or the number of miscounts on the DCT is rather small ($r_{RMT} = .12$, $n = 483$, $p < .05$; $r_{DCT} = -.08$; $n = 479$, $p > .05$). These findings suggest that the magnitude of a positive MCS, in and of itself, is not a good indicator of effort. This makes good sense, when one consid-

Table 5. Mean Memory Comparison Scores (MCS), Rey Memory Test (RMT) Scores, and Dot Counting Test (DCT) Scores by Category of MCS.

			Test Score					
			MCS		RMT		DCT	
Category	MCS range	*n*	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
Strongly Invalid	–5 or lower	17	–6.4	(1.6)	9.0[a]	(2.9)	6.1[b]	(2.6)
Weakly Invalid	–4 to –1	17	–1.3	(0.5)	11.2[a]	(3.2)	3.8[b]	(3.0)
Weakly Valid	0 to 3	17	2.9	(0.2)	13.5	(2.3)	2.2	(1.5)
Strongly Valid	4 or higher	17	9.7	(1.2)	13.4	(2.2)	1.6	(1.6)

Note. Individuals are matched on Auditory Verbal Learning Test (AVLT) trial 1 scores across categories; for AVLT within each category, $M = 4.4$, $SD = 1.4$. Comparison of mean RMT scores revealed significant differences, F (3,64) = 10.7, $p < .05$). [a]These RMT means were significantly different from all other group means. Comparison of DCT mean scores revealed significant differences, F (3, 64) = 13.0, $p < .05$. [b]These DCT means were significantly different from all other group means. Post hoc comparisons were conducted using Fisher's LSD method, overall alpha = .05.

ers that perfect performance on both the WRT and AVLT will result in an MCS = 0. Low magnitude positive MCSs result when performances on WRT and AVLT are both somewhat low or when both are somewhat high.

Table 6 shows a comparison between groups composed of individuals who earned low magnitude positive MCS (MCS = 0 to 3), but who were quite different in WRT and AVLT performance. For the group of 609 defendants, the average WRT total score was 8.2 (*SD* = 3.9) and the average AVLT score was 5.0 (*SD* = 1.8). Individuals in Group 1 earned low magnitude positive MCS as the result of below average WRT (≤ 8) and below average AVLT words recalled first trial (≤ 5; WRT: *M* = 5.9, *SD* = 1.6; AVLT: *M* = 4.0, *SD* = 1.1, *n* = 124). Individuals in Group 2 earned low magnitude positive MCS as the result of above average WRT (≥ 9) and above average AVLT (≥ 6; WRT: *M* = 9.8, *SD* = 1.0; AVLT: *M* = 7.3, *SD* = 1.2, *n* = 50). The average number of RMT items recalled for Group 1 was 11.7 (*SD* = 2.9, *n* = 123); for Group 2 the mean was 14.0 (*SD* = 1.6, *n* = 50; *t* = –5.4, df = 171, *p* < .05, Cohen's *d* = .90). The average number of DCT miscounts for Group 1 was 2.8 (*SD* = 2.4, *n* = 122); for Group 2 the mean was 1.7 (*SD* = 1.4, *n* = 49; *t* = 3.2, df = 169, *p* < .05, Cohen's *d* = 0.53).

DISCUSSION

Greiffenstein et al. (1996) found the normative floor effect to be more sensitive (62% vs. 58%) and more specific (88% vs. 82%) than the personal floor effect in a sample of 150 neuropsychological examinees based on whether they were considered to be truly impaired (*n* = 60) or probably malingering (*n* = 90). Despite these differences, the personal floor effect retains some distinct advantages over the normative floor effect. First, the normative floor effect may be more sensitive to significant impairment than the personal floor effect. In the Greiffenstein et al. (1996) study, 100% of the performances of 5 ''dense amnesiacs'' were classified as invalid by the normative floor effect, but only 40% were classified as such by the personal floor effect. Second, a conclusion of invalidity may be easier to communicate and defend in the presentation of evidence related to competency or other psycholegal issues. That is, an effective counter-argument to detection by the normative floor effect is always that the individual has actual, real, and significant impairment which generated a false positive classification as ''invalid.'' This argument is not tenable for the personal floor effect.

Table 6. Comparison of Performance on RMT and DCT for Individuals with Relatively Low MCS Scores But Large Differences in WRT and AVLT Scores.

			Test Score									
			MCS		RMT		DCT		WRT		AVLT	
Category	MCS range	*n*	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
Group 1	0 to 3	124	1.8	(1.6)	11.6	(2.9)	2.8	(2.4)	5.9	(1.6)	4.0	(1.0)
Group 2	0 to 3	50	2.4	(0.5)	14.0	(1.6)	1.7	(1.4)	9.8	(1.0)	7.3	(1.2)

Note. MCS = Memory comparison score. RMT = Rey Memory Test number items reproduced. DCT = Dot Counting Test number miscounted cards. WRT = Word Recognition Test total score. AVLT = Auditory Verbal Learning Test first trial recall. Participants in Groups 1 and 2 had MCS scores from 0 to 3, but differed in magnitude of WRT and AVLT scores. Group 1 members earned below average WRT and AVLT scores. Group 2 members produced above average WRT and AVLT scores. Differences in RMT scores were significant (*t* = –5.4, *df* = 171, *p* < .05, Cohen's *d* = .90) as were differences in DCT scores (*t* = 3.2, *df* = 169, *p* < .05, Cohen's *d* = .53).

The use of the RMT, DCT, and VIP as comparison tests in these analyses is limited because these tests are not considered to be effective measures of invalid performance of memory capacity. Despite this limitation, significant differences were observed between groups on these measures, and this suggests that the MCS has utility in identifying differences in approach to testing. Future research should investigate the MCS with procedures more sensitive and specific to malingered memory impairment, as well as its incremental validity with respect to such procedures. Future research should also consider the validity of MCS in other populations in which a motivation to feign cognitive and memory impairment exists (e.g., civil litigation, disability determination, and worker's compensation).

Most importantly, other neuropsychological procedures or tests might accommodate the personal floor effect strategy in identifying invalid performances. As noted earlier, the VIP has such a strategy inherent in its consideration of effort; performance errors below the test taker's demonstrated ceiling of ability are examined for their meaning with respect to effort. It may be possible to routinely organize information from neuropsychological testing to maximize opportunities to identify performance below a personal floor.

REFERENCES

Binks, P.G., Gouvier, W.D., & Waters, W.F. (1997). Malingering detection with the Dot Counting Test. *Archives of Clinical Neuropsychology, 12,* 41-46.

Binder, L.M. (1990). Malingering following mild head trauma. *Clinical Neuropsychologist, 4,* 25-36.

Binder, L.M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. *Journal of Clinical and Experimental Neuropsychology, 9,* 167-171.

Borum, R., & Grisso, T. (1995). Psychological test use in criminal forensic evaluations. *Professional Psychology: Research and Practice, 26,* 465-473.

Denney, R.L. (1996). Symptom validity testing of remote memory in a criminal forensic setting. *Archives of Clinical Neuropsychology, 11,* 589-603.

Dusky v United States, 362 U.S. 402 (1960).

Frederick, R.I. (1997). *Validity Indicator Profile manual.* Minnetonka, MN: NSC Assessments.

Frederick, R.I., Carter, M., & Powel, J. (1995). Adapting symptom validity testing to evaluate suspicious complaints of amnesia in medicolegal evaluations. *Bulletin of the American Academy of Psychiatry and the Law, 23,* 231-237.

Frederick, R.I., Crosby, R., & Wynkoop, T. (2000). Performance curve classification of invalid responding on the Validity Indicator Profile. *Archives of Clinical Neuropsychology*, *15*, 281-300.

Frederick, R.I., & Denney, R.L. (1998). Minding your ''*p*s and *q*s'' when conducting forced-choice recognition tests. *The Clinical Neuropsychologist, 12,* 193-205.

Frederick, R.I., Sarfaty, S.D., Johnston, J.D., & Powel, J. (1994). Validation of a detector of response bias on a forced-choice test of nonverbal ability. *Neuropsychology, 8,* 118-125.

Greiffenstein, M.F., Baker, W.J., & Gola, T. (1996). Comparison of multiple scoring methods for Rey's malingered amnesia measures. *Archives of Clinical Neuropsychology, 11*, 283-293.

Guilmette, T.J., Hart, K.J., Giuliano, A.J., & Leininger, B.E. (1994). Detecting simulated memory impairment: Comparison of the Rey Fifteen-Item Test and the Hiscock Forced-Choice Procedure. *The Clinical Neuropsychologist, 8,* 283-294.

Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior, 16,* 257-272.

Hiscock, M., & Hiscock, C.K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology, 11,* 967-974.

Lee, G.P., Loring, D.W., & Martin, R.C. (1992). Rey's 15-Item Memory Test for the detection of malingering: Normative observations on patients with neurological disorders. *Psychological Assessment, 4*, 43-46.

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

Melton, G.B., Petrila, J., Poythress, N.G., & Slobogin, C. (1997). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers* (2nd ed.). New York: Guilford.

Rey, A. (1941). L'examen psychologie dans les cas d'encephalopathie traumatique. *Archives de Psychologie, 28,* 286-340.

Rey, A. (1958). *L'Examen clinique de psychologie.* Paris: Presses Universitaires de France.

Robinson, M.D., & Johnson, J.T. (1996). Recall memory, recognition memory, and the eyewitness confidence-accuracy correlation. *Journal of Applied Psychology, 81,* 587-594.

Rogers, R., Bagby, R.M., & Dickens, S.E. (1992). *Structured Interview of Reported Symptoms (SIRS) and professional manual.* Odessa, FL.: Psychological Assessment Resources.

Rogers, R., Harrell, E.H., & Liff, C.D. (1993). Feigning neuropsychological impairment: A critical re-

view of methodological and clinical considerations. *Clinical Psychology Review, 13,* 255-274.
Schretlen, D., Brandt, J., Krafft, L., & Van Gorp, W. (1991). Some caveats in using the Rey 15-Item Memory Test to detect malingered amnesia. *Psychological Assessment, 3,* 667-672.
Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.
Slick, D., Hopp, G., Strauss, E., Hunter, M., & Pinch, D. (1994). Detecting dissimulation: Profiles of simulated malingerers, traumatic brain-injury patients, and normal controls on a revised version of Hiscock and Hiscock's forced-choice memory test. *Journal of Clinical and Experimental Neuropsychology, 16,* 472-481.
Tombaugh, T.N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment, 9,* 260-268.