# A Review of Rey's Strategies for Detecting Malingered Neuropsychological Impairment

Richard I. Frederick

*U.S. Medical Center for Federal Prisoners*

**SUMMARY.** André Rey was a pioneer in clinical psychology, child clinical psychology, and neuropsychology. His contributions to the understanding of brain-behavior relationships were novel, creative, and highly regarded. Rey developed many eponymous tests and procedures. Three are relatively well known in American neuropsychological literature as "tests of malingering": The Rey 15-Item Memory Test (RMT), the Rey Word Recognition Test (WRT), and the Rey Dot Counting Test (DCT), referred to collectively in this paper as the "Rey malingering tests." By citing his original work, this article reports Rey's method of assessing the validity of clinical presentation, his instructional sets for each test, and his process of interpreting test results. Additionally, this

Richard I. Frederick is affiliated with the U.S. Medical Center for Federal Prisoners, Springfield, MO.

Address correspondence to: Richard I. Frederick, PhD, Department of Psychology, U.S. Medical Center for Federal Prisoners, Springfield, MO 65807 (E-mail: richardfrederick@yahoo.com).

The author thanks Teresinha Rey and André-Louis Rey for their generous assistance and hospitality in discussing André Rey's work. Some information about Rey's personal history comes from discussions they and the author had on May 21 and 22, 1998. The translations and photograph in the article are used with their permission. The author also thanks Carlos Tomelleri and Thomas Fullerton for their translations of Rey's writings.

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

[Haworth co-indexing entry note]: "A Review of Rey's Strategies for Detecting Malingered Neuropsychological Impairment." Frederick, Richard I. Co-published simultaneously in *Journal of Forensic Neuropsychology* (The Haworth Medical Press, an imprint of The Haworth Press, Inc.) Vol. 2, No. 3/4, 2002, pp. 1-25; and: *Detection of Response Bias in Forensic Neuropsychology* (ed. Jim Hom, and Robert L. Denney) The Haworth Medical Press, an imprint of The Haworth Press, Inc., 2002, pp. 1-25. Single or multiple copies of this article are available for a fee from The Haworth Document Delivery Service [1-800-HAWORTH, 9:00 a.m. - 5.00 p.m. (EST). E-mail address: getinfo@haworthpressinc.com].

article reviews published research regarding the efficacy of the Rey malingering tests as they have been used in this country. Most instructional sets and interpretive strategies generally have not followed Rey, but have followed those represented as Rey's by Lezak (1983, 1995). 

**KEYWORDS.** Malingering, neuropsychological assessment, psychological evidence, André Rey

André Rey at the International Congress of Psychology, Rome, Italy, 1958. Photo courtesy of Teresinha Rey.

André Rey was born in 1906 in Lausanne, Switzerland. His first published paper concerned recent developments in French psychology and philosophy (Rey, 1928). In 1929, he went to Geneva to work with Edouard Claparède, a physician and psychologist, who founded the Institute of Educational Sciences and initiated clinical psychology in Switzerland (Kaschel, Goldenberg, Goldstein, Risberg, Laaksonen, Gaillard, Perret, Tissot, & Oktem-Tanor, 1994). The Institute was linked to the clinical psychology laboratory in the neurology service at Geneva Hospital (l'Hôpital Cantoral de Genève). The objectives of the laboratory were to identify the extent of cognitive deterioration in traumatic brain injury and alcoholism (Kaschel et al., 1994). In 1934, Rey completed his doctoral dissertation, which concerned the assessment of practical intelligence of children. His experiments evaluated the methods by which children from ages 4 to 8 could solve practical problems. For example, he studied how children could obtain by indirect means a piece of candy suspended by a string (Lambercier & Rey, 1935; Winstch, 1935). In 1936, he received a Rockefeller scholarship and went to Harvard University to work with Karl Lashley for two years. His work included im-

proving some of Lashley's methodology for experimentation with animals (Rey, 1938).

Upon his return to Geneva, he resumed his work at the Institute. The problems he addressed through clinical practice were methods of identifying subtle neuropsychological signs, neuroreception, muscular strength and tonus, constructional praxia, visual perception, articulation of language, motor coordination, balance, memory and learning, and concentration. He worked with developmentally delayed children and with patients with different sorts of brain pathology. When Claparède died in 1940, Rey assumed the position as director of the psychology laboratory. He was professor of applied psychology in the Institute of Educational Sciences from 1949 onwards.

In 1954 and 1955, he studied the adaptability of Jewish children from North Africa to live in Palestine, developing language-free general intellectual measures (Rey, 1957). In 1956, he went to Brazil at the request of its government to develop a psychology curriculum at Belo Horizonte. During his stay there, he met his future wife, Teresinha, who went a little later to Geneva for further studies and worked as a psychologist, evaluating and treating children with physical and mental disabilities in partnership with Rey. In 1958, he was named Lecturer for Clinical Psychology for the Faculty of Medicine at Geneva Hospital. In 1958, he began a continuing visiting professorship of psychology in Lyons, France. He had a full private clinical practice from 1938 until 1965. Rey was an advanced amateur entomologist and botanist (he served as president of the Society of Physical and Natural History in Geneva) and published works on the behavior of spiders and caterpillars.

From 1928 until his death in 1965, he published more than 170 articles and books concerning animal psychology, experimental psychology, clinical techniques, test development, clinical interviewing, and diagnostic tools. His works include investigations of elementary brain functions (development of perception: visual, tactual, gustatory; motor coordination, spatial localization, sensory-motor accommodation), more complex brain functions (mental imagery, development of mental processes, and attention), attention and learning (acquisition of learning, associative learning, comparison of learning curves between infants and animals), developmental processes (children's abilities, child psychopathology, developmental delays), assessment techniques (personality tests, intellectual assessment, developmental drawing tests, projective drawing tests), effects of trauma on learning, traumatic brain injury and recovery, and professional issues (relationships between psychologists and psychiatrists, education and training of psychologists). His research

was typified by analyses of large data bases collected from clinical practice.

In 1967, the Association of the Friends of André Rey (including Georges de Morsier, Paul A. Osterrieth, Robert Dottrens, and Marc Richelle) published a memorial volume: the *Hommage a André Rey*. A group composed of his last assistants, working with Teresinha Rey, published five volumes of previously unpublished tests. Additionally, Teresinha Rey has been organizing and cataloging Rey's works for the University of Geneva, where they will remain available for future researchers.

Rey is perhaps best known in American psychology as the developer of the Rey Auditory Verbal Learning Test (AVLT), the Rey-Osterrieth Complex Figure Test, and for several techniques for assessing feigned neuropsychological impairment (Lezak, 1995). Lezak also noted his skin writing techniques, and Taylor (1959) reported his work concerning assessment of children's maturational skills by human figure drawings (Rey, 1946). Although his contributions are well-known throughout Europe, only a small body of his work is known in the United States. This article reviews Rey's work particularly with respect to the evaluation of feigned neuropsychological impairment.

## *GENERAL CAVEATS ON MALINGERING DETECTION BY REY*

Rey developed interest in malingering detection partly in response to accusations by insurance companies that most, if not all, claimants were probably only feigning impairment. His techniques were not limited to the malingering tests described in this paper. He developed many techniques and procedures, most of which depended primarily upon a performance curve analysis; that is, a comparison of achievements across a range of test difficulties (see Frederick, Crosby, & Wynkoop, 2000; Rogers, Harrell, & Liff, 1993). Neurologist Georges de Morsier, wrote:

> . . . Rey demonstrated that all simulators (the fear of the insurance experts) were immediately discovered by the characteristic nature of the curves. From the initiation of regular collaboration with the clinic (1940), Rey continually perfected and improved these methods of assessing memory and perception, thus allowing one to see

> in a glance from its graphical representation the mental capacity of the patient and the deficiencies acquired since the patient's illness or trauma. (excerpted from *Hommage a André Rey*, 1967; p. 19)

Rey's general approach to assessing malingering included an understanding that some idiosyncratic presentations could mimic malingering, that use of a single sign was insufficient to make effective discriminations between individuals with real impairment and malingerers, and that an understanding of the context of the evaluation and the potential motivations of malingerers was paramount:

> In this case it is a matter of unmasking the individual, remembering that the exaggeration of symptoms may constitute in itself a behavioral disorder and not always a calculated behavior. It is important, first of all, to evaluate the cultural level of the patient, then to characterize his balance of satisfactions and life style, information which often permits us to understand the need or self-interest leading to the patient's exaggeration or simulation. (Rey, 1958; p. 121)

## *REY FIFTEEN-ITEM MEMORY TEST*

The RMT is probably the best known of Rey's procedures. It generally became known in American neuropsychological literature through Muriel Lezak's second edition of *Neuropsychological Assessment* (1983). The first experimental data for the RMT were reported by Goldberg and Miller (1986). Since that time, a large number of articles have appeared, most of which highlight the limitations of the test in distinguishing between malingerers and individuals with significant neuropsychological impairment.

### *Instructions*

Rey's instructions for this test are as follows:

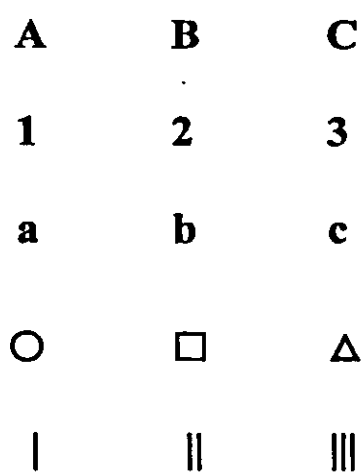
> First of all we ask him to freely express the data retained after listening to a series of 15 words. Let us suppose this immediate response is very weak (3 or 4 words). Next we will present the following 15 signs grouped by three. [Note: see Figure 1]

> We expose them for 30 seconds, then ask the subject to reproduce them without worrying about their order. We insist upon the fact that there are 15 different elements. During the duration of the exposure we state each series so as to rule out any ambiguity: "Capital A, B, C; 1, 2, 3; a, b, c, small letters; a circle, a square, a triangle; one stick, two sticks, three sticks." We allow two minutes for reproducing the data viewed. (Rey, 1958; p. 121-122)

Lezak (1983) modified these instructions substantially, but made no references to administering the test immediately following a 15-item word list (ostensibly for contrast) or, more particularly, following a weak performance on such a test. She identified the period of memorization as only 10 seconds, and added "a 10-or 15-second quiet delay period can be interpolated" (p. 619). The instructional set of Goldberg and Miller (1986) has been the basis for most research since then: "I am going to show you a card with 15 [emphasized] things on it to remember. When I take the card away, I want you to write down as many of the 15 things as you can remember" (p. 794). After 5 seconds, a prompt was given: "Be sure to look at all of them." The stimulus items were removed after a 10-second exposure.

Other researchers have emphasized different aspects of the instructions. Schretlen, Brandt, Krafft, and van Gorp (1991) and Guilmette,

FIGURE 1



Hart, Giuliano, and Leininger (1994) emphasized there were 15 different items to remember. Arnett, Hammeke, and Schwartz (1995) presented the RMT as "a very difficult memory test" and added "this is a hard [emphasized] test."

Rogers et al. (1993) argued it might be unethical and clinically unnecessary to describe the task as "difficult." Similarly, Bernard and Fowler (1990) considered it counterproductive to tell patients a test was difficult when it is obviously designed to promote recall. These points are well made. It seems clear from Rey's original instructions that he used the test when performance on a reasonably difficult task reflected evidence of poor effort, and he "insisted" there were 15 different elements only to induce the dissimulating patient to relate the task to the previous test. Otherwise, there does not appear to have been any attempt by Rey to make the task seem difficult by instructional set. In research simulation and clinical practice, Frederick, Sarfaty, Johnston, and Powel (1994), Frederick (1997), and Frederick (2000b) maintained the primary instructional set of Lezak (1983); that is, a 10-second exposure and a 10-second interpolated delay, but administered the Rey after the completion of the AVLT, beginning with the transitional statement: "Now I'm going to *show* you 15 things to remember."

### *Scoring*

Rey was interested in the number of items completed and the number of groups completed. Most research on the Rey has focused on counting the number of completed items. Griffin, Normington, and Glassmire (1996) reported a qualitative analysis of the RMT, focusing on whether the rows were completed out of sequence, when characters were modified (e.g., a reversal of "b" to "d") or embellished (e.g., when the sticks were transformed into Roman numerals).

### *Interpretation*

Rey's interpretation follows:

> If the patient is not intellectually weak, if he is not suffering serious mental deterioration, and especially if he is sincere and cooperates faithfully, he will reproduce the 15 signs without difficulty. In fact, there are only 5 data to be retained at the most, automatisms and associations assuring the reproduction of the 10 others. One may tolerate the forgetting of a group, most often the

> sticks and one of the geometric figures. Forgetting two groups is already suspect and forgetting three allows the conclusion of exaggeration and doubtful collaboration by the subject whose intellectual level is normal and who, moreover, takes great care to describe his disorders. When the 15 signs are reproduced correctly, and the patient has remarked, often spontaneously, that the test was easy, or much easier than the preceding one, we will be able to establish in his performance a collaboration sufficient, at least, for us to push further the investigations and we shall be allowed to take into consideration the difficulties presented on the test of the immediate expression of 15 words.
>
> Let us note, moreover, that if on this first test, the result had been normal, it would be in itself an indication of sincerity and of collaboration, and one could immediately concede a certain amount of credit to the complaints expressed, and direct research onto aspects of memory more differentiated than a simple, immediate evocation of words. . . . One will eliminate without difficulty the cases of intellectual debility or of easily identifiable insanity to other signs and results. (Rey, 1958; pp. 122-123)

Despite Rey's caveats, Lezak (1983) reported only, "Anyone who is not significantly deteriorated can recall at least three of the five character sets" (p. 619). This cut-off score (i.e., fewer than 9 items reproduced) was supported by Goldberg and Miller (1986) and Bernard and Fowler (1990), who considered any performance less than nine items predictive of malingering. Frederick et al. (1994) found this cut-off had a 40% sensitivity and 100% specificity for a large sample of simulating malingerers and normal controls. Greiffenstein, Baker, and Gola (1994) reported that cut-off scores of 9 or 10 had sensitivity of 62% for probable malingerers and specificities of 88% and 81% for individuals with bona fide traumatic brain injury.

The cut-off has often been criticized as too nonspecific for persons with true impairment (Hays, Emmons, & Lawson, 1993; Lee, Loring, & Martin, 1992; Schretlen et al., 1991). Lee et al. (1992) recommended using a cut-off score of fewer than 8 items reproduced, despite only a small gain in specificity for doing so (95% instead of 93%). Hays et al. (1993) recommended interpretation of RMT scores with respect to measured IQ levels, apparently failing to consider the possibility participants had feigned their IQ test performance. Others have considered even the cut-off score of 8 to be too nonspecific (Arnett et al., 1995; Guilmette et al., 1994). Arnett et al. (1995) recommended a cut-off

score of fewer than two complete rows. Simon (1994) recommended against the use of less than three rows, finding it too nonspecific (57%) for his control group of severely ill insanity acquittees, although Back, Boone, Edwards, Parks, Burgoyne, and Silver (1996) found the false positive rate (using a cut-off of fewer than 9 items) for 30 patients with schizophrenia was only 13%.

Other researchers have recommended higher cut-off scores, concluding a cut-off of less than 9 was too insensitive (Bernard, Houston, & Natoli, 1993). Greiffenstein, Baker, and Gola (1996) reported that a cut-off score of fewer than 10 items maintained both high specificity and sensitivity. DiCarlo, Gfeller, and Drury (1996) recommend a cutting score of fewer than 14 items(!). Iverson and Franzen (1996; see also Arnett & Franzen, 1997) modified the test to a 16-item test, eliminating the geometric figures, hoping to make the test more sensitive to malingering. The procedure has also been modified by Griffin, Glassmire, Henderson, and McCann (1997) in an attempt to increase sensitivity and specificity. Frederick (2000b) compared a number of the reported sensitivities and specificities for many of these studies.

Rey would seem to readily concede that the RMT is likely to be sensitive to significant mental, neuropsychological, or intellectual impairment. His own caveats include: "*If the patient is not intellectually weak, if he is not suffering serious mental deterioration . . .*" and "Forgetting two groups is already suspect and forgetting three allows the conclusion of exaggeration and doubtful collaboration by the subject *whose intellectual level is normal* and who, moreover, takes great care to describe his disorders" (emphasis added). In support of these contentions, Frederick (2000b) used a variety of methods to estimate that the cut-off of fewer than 9 items was moderately sensitive (58% to 89%) and highly specific (96% to 98%) in a criminal forensic setting with a reasonably low base rate of significant neuropsychological injury or impairment. Analysis using receiver operating characteristic curves for the sample of 723 criminal defendants yielded values for the area under the curve of 0.94 to 0.98, meaning that randomly drawing an RMT score from a pure sample of malingerers would be lower (i.e., fewer items reproduced) than a randomly drawn RMT score from a pure sample of non-malingering criminal defendants 94% to 98% of the time.

Rey's own description of the RMT is as "a very simple maneuver," using it as an example in his discussion perhaps only to avoid detailed descriptions of more elaborate procedures whose "details . . . it is preferable not to disclose."[1] (Indeed, Frederick et al., 1994, found the RMT's sensitivity dropped to 5% when simulating malingerers re-

ceived only minor coaching.) I suspect Rey would be astonished to find his "simple maneuver" had been elevated to the status of a primary detection strategy and then criticized for its presumptuousness.

## *REY DOT COUNTING TEST*

The DCT and WRT have been evaluated far less than the RMT, but they were his earliest examples of applying malingering detection methodology. The following passage regarding the DCT is from a journal article published in 1941:

> Investigating the disordered character of responses:
>
> Tests to determine sensory thresholds can be utilized for this purpose. Without the subject being able to notice, the intensity of the stimulus is varied in an irregular fashion. The malingerer, sometimes some psychopaths, furnish responses which do not exhibit any relationship between the intensity of the stimulus and that of the perception. For example, determination of tactile threshold with the Weber compass, a study of auditive perception with the acumetric inductor of Foy, etc. The same principle can be applied to more complex psychological operations. For instance, we will present to a subject, in an irregular order, tasks of increasing difficulty. If the times of execution are proportional to the difficulty of the tasks, we are in the presence of a normal or abnormal performance depending on the magnitude of these execution times and in such case, this is not suspect. If this progression fails to occur, if we register gross irregularities, the performance is suspect, or at least it poses a problem that needs to be elucidated.
>
> An example of a test of this type follows: Six cards are presented successively to the subject, each one containing a certain number of dots distributed irregularly. The subject has to count the dots as rapidly as possible without using his finger. The time that transpires between the instant of the presentation and the response is noted.
>
> The cards, numbered one through six, contain respectively seven, 11, 15, 19, 23 and 27 dots. (Each dot is one square millimeter and they are equidistant by .5 centimeters.) The cards are presented in the following order: two, four, three, five, six, one. The establishment of each total (the dot count) necessitates a certain effort of attention that will require a longer time in proportion to the

number of dots counted. If the subject works in earnest we will find a progression of times as we arrange the responses according to the increasing number of dots.

The following table allows one to form an exact idea of the speed of the responses and the progression of times among adult normal individuals:

| Percentile: | 100 | 75 | 50 | 25 | 0 |
|---|---|---|---|---|---|
| Card 1: 7 dots | 1 sec | 2 sec | 4 sec | 5 sec | 11 sec |
| 2: 11 | 2 | 3 | 4 | 5 | 17 |
| 3: 15 | 3 | 4 | 6 | 7 | 17 |
| 4: 19 | 4 | 6 | 7 | 9 | 19 |
| 5: 23 | 5 | 8 | 10 | 12 | 30 |
| 6: 27 | 6 | 9 | 11 | 16 | 30 |

The 50th percentile column shows typical results of average individuals, percentiles 100 and 75 columns of fast individuals, and the 25 percentile column those of slow individuals. The zero column displays pathological results originating in subjects who suffer from traumatic encephalopathy (sequelae of head injury).

The examination of these figures allows interesting determinations.

A. The progression is more sharp and regular when we are dealing with faster individuals who have made an effort to perform well.

B. The progression is less regular among individuals who are slower or who suffer from attention deficits. This however is a relative irregularity for which our table furnishes the tolerance limit. If the irregularity would become more pronounced it would give rise to question about the good will and good faith of the subject. We can, however, tolerate a strongly aberrant result in the midst of the progression. Actually, an accident can happen, or a scrupulous individual may repeat the count of the groups of dots. One must take into consideration errors whose distribution corresponds to indications furnished by the time measures. An error in this account is less suspect if it occurs in a group where the dots are more numerous. Finally, in the presence of suspect performances, it is always useful to repeat these tests several times.

It needs to be added that the principle of the multi-sense character of a single test needs to be applied to the tracking down of malingering as it applies to psychological diagnosis in general. In the psychological plane, a single suspect result raises a hypothesis and orients research; it does not create a certainty.

Use of tests of different degrees of difficulty combined with changes of attitude of the examiner.

This is the group of procedures that is most important and most efficacious to investigate malingering and absence of effort. Let us suppose the psychologist has available several tests of the same type for which it has been established that differences in performance are the rule. In that manner, Test B must produce among all conscientious individuals quantitatively superior results (timewise) compared to Test A.

We will begin by presenting the easy test (A) after patiently having listened to the patient's complaints. We will let him work freely and we will observe him with benevolence. If the result is poor and there is otherwise reason to suspect the good will of the subject we will then proceed to Test B, which is the more difficult one. At this moment we will change our attitude. We will declare to the patient we are afraid he has abused the benevolence that we have shown him, that the result he has furnished is suspect, and that from this point on we expect a more conscientious performance from him. We then give him Test B. If, with this more difficult test, his performance is superior or equal to that produced on the easier test then this is an indication that confirms the presence of an initial tendency to not make much of an effort.

We have already proposed a great number of tests that allow us to realize the maneuver we have just described. For example, here are two procedures which are perhaps new and will augment a stock of useful procedures. Let us return to the test of groups of dots. It will serve as Test B difficult. Let's imagine presently a Test A of the same type but much easier. Instead of using irregular groups, we are going to distribute the dots in geometric groups. The total number will be much easier to establish and consequently responses will be faster. On six cards we will distribute groups numbering respectively 8, 12, 16, 20, 24, and 28 dots. Each group will be composed of the following groups–8 would be two squares of four, 12 would be two squares of five plus two separate dots, 16 would be four diamonds of four, 20 would be four groups

of five (square of four dots with one dot in the middle), 24 would be four groups of six (rectangle of six dots), 28 would be four groups of five plus two groups of four. The cards numbered from one to six are presented in the following order: 2, 4, 3, 5, 6, 1. This test elicits a psychological mechanism which is different from that of the first test. To obtain the total number of dots in the irregular groupings it is necessary to count dots one by one; attention plays a principal role. Whenever the dots are grouped the numerical value of each group is perceived almost instantaneously and it is sufficient to make an elementary addition or multiplication to obtain the total. (In giving instructions one must insist that the subject count as fast as possible and using the quickest means.) The number of dots in this case does not play more than an insignificant role and the progression of times of response is very weak. Following is a table that gives an idea of the rapidity of response among normal adults. We will notice the weak dispersion of the results.

| Percentile: | | 100 | 75 | 50 | 25 | 0 |
|---|---|---|---|---|---|---|
| Card 1: | 8 dots | .5 sec | 1 sec | 1 sec | 2 sec | 3 sec |
| | 12 | 1 | 2 | 2 | 2 | 3 |
| | 16 | 1 | 2 | 2 | 4 | 5 |
| | 20 | 1 | 1 | 2 | 4 | 5 |
| | 24 | 2 | 2 | 2 | 5 | 6 |
| | 28 | 2 | 2 | 3 | 5 | 7 |

Test A is clearly easier than the previous test (B) and we should have a total time for (A) which is less than the total time for (B). We can then utilize the system of test AB according to the procedure described above. These are examples of normal reaction.

Adult Subjects (comparison of total times for the six cards of each series)

| | Test A (Easy) Grouped dots | Test B (Difficult) Irregularly distributed dots | Difference B − A |
|---|---|---|---|
| Subject X | 9 sec | 31 sec | +22 |
| Subject Y | 11 | 31 | +20 |
| Subject Z | 12 | 33 | +21 |

The following are examples of pathological reactions or responses that do not elicit suspicion:

Traumatic Encephalopathy, chronic case, the accident happened approximately two years ago.

| | A | B | Difference B − A |
|---|---|---|---|
| Subject L | 21 sec | 37 sec | +16 |
| Subject M | 22 | 38 | +16 |
| Subject N | 24 | 69 | +45 |

In these three cases, the slowness of response is explained by slowness of mentation and troubles of attention and perception. It is proportional to the degree of difficulty of the tests.

Here is an example of a suspect result:

| | A | B | Difference B − A |
|---|---|---|---|
| Subject D | 57 sec | 42 sec | −15 |

The subject spends more time with the easy test than with the difficult test. He has counted the dots of the figures in test A instead of evaluating the groups and proceeding arithmetically. This is a case of idiopathic feeble-mindedness (Rey, 1941; pp. 312-316).

### *Instructions*

In this passage, Rey described two separate procedures using the dots. One uses the "irregularly distributed" dots or "ungrouped" dots, as they are commonly referred to. Rey directed patients to count the dots as quickly as possible by the fastest means possible. The other procedure involves both sets of dots, but the grouped dots are presented first, and a "shift of attitude" constituted a principal aspect of the administration. Lezak (1983) reported the procedure as either an ungrouped administration or one in which grouped dots were presented after the ungrouped dots. The combined presentation with no shift of attitude seems to have been the standard form of administration in research since then (Binks, Gouvier, & Waters, 1997; Frederick et al., 1994; Lee, Boone, et al., 2000; Paul, Franzen, Cohen, & Fremouw, 1992; Rose, Hall, & Szalda-Petree, 1998), but Frederick (1997) included instruc-

tions to not count out loud or by using one's fingers (to minimize counting strategies that would likely slow down the speed of counting).

### *Scoring*

Frederick et al. (1994) used only a comparison of counting times for the grouped and ungrouped dots, identifying positive findings as any occurrence of ungrouped counting time faster than grouped counting time. Paul et al. (1992) and Lee et al. (2000) also scored the number of counting errors for ungrouped and grouped dots, and for all dots. Rose et al. (1998) and Binks et al. (1997) included counts of the number of deviations from linearity for ungrouped and grouped sets of dots. Frederick (1997) proposed a scoring method which transformed information about counting times and counting errors (total of absolute errors from correct number of dots per card) into a single score.

### *Interpretation*

The DCT has not generally been found to be very sensitive to the presence of malingering. Frederick et al. (1994) found the presence of faster time to count ungrouped dots had only a 5% sensitivity among simulating malingerers. Rose et al. (1998) did not find any useful aspect of the DCT in the detection of malingering. Paul et al. (1992) generated a number of measures to discriminate successfully between simulators and controls, but they did not cross validate their cut-off scores. Binks et al. (1997) determined that the score which best discriminated between simulators and non-simulators was the total number of counting errors across all 12 cards, but no cut-off score was reported.

Rey's approach to the DCT, *performance curve analysis*, has been described by Rogers et al. (1993) and investigated by Frederick and Foster (1991), Frederick et al. (1994), Frederick (1997), Frederick and Crosby (2000), Frederick, Crosby, and Wykoop (2000), Gudjonsson and Shackleton (1986), and McKinzey, Podd, Krehbel, and Raven (1999). It is likely performance curve analysis has proven inadequate for DCT only because there are so few instances to compare. As a consequence, single arithmetic errors or recounting by the diligent patient significantly disrupt the linearity of the performance curve. Other tests which incorporate a performance curve analysis strategy, such as the Validity Indicator Profile (VIP; see Frederick, Crosby, & Wynkoop, 2000), generate the performance curve by smoothing a large number of

data points (i.e., by "hanning"), so that single instances of error do not substantially alter the general shape of the performance curve.

## *REY WORD RECOGNITION TEST*

The previous passage continues with a discussion of how the procedure of comparing easy and difficult tasks applies to the WRT:

> Let's go on to another example. It refers to a procedure which exploits the normal correlations between evocative memory and recognition memory. The easy test (A) will consist of recognizing words which are heard mixed with unknown words. We show the subject the five columns of words in the following table. We turn the page before he has been able to examine them one-by-one. We announce to him that we are going to read a certain number of words and that these words will be found in the columns shown previously, mixed with other unknown words. Instructions finished, the subject must take the paper and underline all the words that he remembers.
>
> Series of 30 words containing 15 words to recognize:

| | | | | |
|---|---|---|---|---|
| goodbye | tomorrow | tub | herb | lightning |
| half | toy | grain | boot | knee |
| gift | wool | pike | ash | hair |
| chamois | force | entrance | local | fly |
| morning | fashion | concert | filet | fault |
| airplane | shoe polish | biscuit | corridor | countryman |

> Series of 15 words to read (one per second): *half, chamois, morning, toy, wool, shoe polish, grain, pike, herb, ash, local, corridor, knee, fault, countryman.*
>
> After the reading we introduce a pause of five seconds and then we present the page carrying the five columns. We allow two minutes for the subject to underline the words. If the result is poor and if there is reason otherwise to suspect or to distrust, we will proceed to test B which is more difficult. It is a good idea to introduce a pause of ten minutes between the two tests. Before starting test B, we will have emphasized the poor performance on A and required more conscientious effort on the new work.

Test B consists of recalling as many words as possible after one hearing of the series. Reading series is as follows: *drum, curtain, belt, coffee, school, relative, sun, garden, cap, countryman, mustache, turkey, color, house, stream.* We introduce a pause of five seconds after the last word is read. The free recall period lasts two minutes. The following is a table of the results:

| Percentile: | 100 | 75 | 50 | 25 | 0 |
|---|---|---|---|---|---|
| *Recall:* | | | | | |
| Educated Adult Subjects (high school education) | 14 | 10 | 8 | 7 | 5 |
| Less Educated Subjects (elementary education) | 10 | 7 | 6 | 4 | 2 |
| *Recognition:* | | | | | |
| Educated | 15 | 15 | 14 | 12 | 8 |
| Less educated | 14 | 12 | 10 | 7 | 5 |

If we examine these figures, we realize that in general more words are recognized than recalled whatever the intrinsic value of memory may be. Just suppose that the patient at the time of test A has not recognized more than three words which is a clearly pathological result, but he is able to recall four during test B, which is more difficult. Let's suppose that all question of incomprehension or technical error is eliminated, that there is no alexia and that there has not been any transient malaise or discomfort. If all of these conditions are fulfilled, we are in the presence either of an individual of a particular mnestic type which would be quite rare, or an individual who does not make an effort or who is trying to fool us. In the first case, the type must remain constant and we will reencounter it in repeating the test in an equivalent form and in pushing further the comparative study of mechanisms of recall and recognition. In the second case, retesting will not put in evidence any systematic difference; we shall be increasingly diffident, and very soon we will be able to attempt to put the subject in the presence of facts and figures. His reaction will then be quite interesting (Rey, 1941; pp. 316-318).

### *Instructions*

Lezak (1983, 1995) and Greiffenstein et al. (1994) followed the instructional set in generally the same fashion as Rey; that is, the 15-word list is presented for memorization and then a 30-word list is presented on a sheet of paper. The patient circles the words which were presented for memorization. Frederick et al. (1994), Frederick (1997), Frederick (2000a), and Greiffenstein et al. (1996) modified the instructional set by eliminating the presentation of the sheet of words, instead presenting the 30-word recognition list orally, directing participants to quickly determine if each word had been on the 15-word memorization list (a two-alternative forced-choice technique; see Frederick, 1997, for a detailed description of accompanying instructions and practice items).

### *Scoring and Interpretation*

Generally, scoring has proceeded in the same way as described by Rey: a comparison of total recognized words to total recalled words (on the first trial). Frederick et al. (1994) found this method was about 4% sensitive and about 97% specific (but Greiffenstein et al., 1996, reported a much higher sensitivity of 58%, with specificity of 88%). Greiffenstein et al. (1996) examined various methods of deriving a useful cut-off score; the preferred score is calculated by adding correct recognitions to correct rejections. This *accuracy score* demonstrated a sensitivity of 80% and specificity of 85% in discriminating between probable malingerers and patients with traumatic brain injury. Frederick (2000a) computed a *memory comparison score* which was derived by first totaling the number of correctly recognized words and then subtracting the number of incorrectly recognized words, and then by subtracting the number of correctly recalled words (out of 15) from the first trial of the AVLT. Frederick (2000a) reported that values greater than or equal to 0 were consistent with other evidence of compliance, and values less than 0 were strongly associated with other indications of malingering.

## *DAUBERT ISSUES*

*RMT.* I agree with Vallabhajosula and van Gorp (2001) that the RMT does not enjoy general acceptance among neuropsychologists or forensic psychologists as a *primary detection strategy*. Published error rates

for various cut-off scores for the RMT vary greatly, and probably reflect the limitations of research design (Frederick, 2000b). Although I do not believe the RMT meets *Daubert* criteria for acceptability as evidence as a primary detection strategy, I certainly see no problem with including the test as part of an overall strategy to evaluate the believability of performance in the way described by Rey.

*DCT.* I conclude the DCT does not meet *Daubert* criteria for acceptability as evidence regarding malingered cognitive impairment. It does not enjoy general acceptance as a procedure to investigate malingering, and it does not have a research literature supporting any of the potential scoring routines which have been investigated for it.

*WRT.* I do believe the process of evaluating memory disorders by comparing recall and recognition memory (e.g., by use of the WRT and AVLT) meets *Daubert* criteria. The processes described (by Frederick, 2000; Greiffenstein et al., 1994, 1996) in peer-reviewed journals consistently show good sensitivity and specificity (see also Haines & Norris, 2001, for differences among patient groups and different types of malingering groups). The problem with comparison of recognition and recall memory performances tends to be that the process is not sensitive to malingering, but accounting for false positive recognitions seems to raise the sensitivity substantially, with only a minimal decrement in specificity. The practice is logically coherent, and probably has enjoyed general acceptance as a strategy since Lezak's description of the process in 1983, if not earlier.

## *USING THE REY MALINGERING TESTS*

### *Using the Tests for Research Purposes*

I have used the Rey tests to conduct research on malingering detection for the past 10 years. I have constructed criterion groups based on Rey test performance (e.g., Frederick, 1997). Given that most of the clinical patients I have sampled to form criterion groups did not have severe neuropsychological impairment or mental retardation, it seemed likely that positive scores on two or three Rey malingering tests were predictive of malingering, given the generally high specificity of the tests. On the other hand, given the generally low estimates of sensitivity for the tests, it seemed likely that "negative" criterion groups (inclusion rule: "no positive scores" on the tests) would probably be significantly contaminated by malingerers who were not so obvious as to perform too

badly on the Rey tests. A consequence of this is that the test or technique being validated will appear to have lower specificity than it actually does. I have discussed this problem at length in Frederick (2000b) and argued that the RMT probably has excellent discriminatory power in a forensic criminal setting in which there is a low base rate of significant neuropsychological impairment.

Currently, my preference is to use the Rey malingering tests as dependent variables. Not only does this forestall any criticism for using them to establish criterion groups, I have found they reliably demonstrate significant mean separations (e.g., Cohen's $d \geq 1.0$) for criterion groups that appear to be substantially different with respect to an intention to perform well on collaterally administered cognitive tests (Frederick, 2000b).

### *Using the Tests Clinically*

I would encourage any forensic evaluator to routinely administer the RMT in the manner Rey described (i.e., after a 15-word recall task), but perhaps using a modified form of Lezak's instructional set, for example, as suggested by Frederick (1997). Just as an illustration, consider the data in Table 1. These data are from 686 criminal defendants who completed the AVLT, the RMT, and the VIP. I have previously described the sample from which they are taken in Frederick, Crosby, and Wynkoop (2000), and Frederick (2000a, 2000b). Classifications of Irrelevant and Malingering on the VIP reflect an intention to respond incor-

TABLE 1. Cross-Classification of Categorizations on the Validity Indicator Profile (VIP), the Rey 15-Item Memory Test (RMT), and the Rey Auditory Verbal Learning Test (AVLT)

| | | | VIP Classification | | | |
|---|---|---|---|---|---|---|
| | | | Motivated to Perform Well | | Motivated to Perform Poorly | |
| AVLT | RMT | *n* | Compliant | Careless | Irrelevant | Malingering |
| > 4 items | > 8 items | 395 | 70.6% | 24.9% | 6.1% | 0.5% |
| < 5 items | > 8 items | 224 | 51.3% | 38 8% | 9.4% | 0.4% |
| > 4 items | < 9 items | 14 | 0.0% | 50 0% | 35.7% | 14 3% |
| < 5 items | < 9 items | 53 | 13.2% | 28.3% | 45.3% | 13.2% |

*Note* $N = 686$ Partcipants are all criminal defendants. AVLT is the number of words recalled correctly on the first trial. RMT is the number of items reproduce. VIP classifications refer to its best classification of approach to testing without any additional contextual information. The classifications here refer to performance on the VIP nonverbal subtest only.

rectly. These data support Rey's view that the RMT might amplify the meaning of poor performance on the AVLT. It is easily observed that the rate at which defendants are classified as "motivated to perform poorly," is correlated with their joint performance on the 15-word recall list and the RMT. Most notably, when "weak" performances on the first trial of the word-recall task are followed by acceptable performances on the RMT, the rate at which defendants are classified as motivated to perform poorly is only 9.8% (and 51.3% are classified as Compliant), but when "weak" performances are followed by unacceptable performances on the RMT, the rate of classification as motivated to perform poorly is 58.5%.

In this light, I routinely tell others that I find the RMT to be one of the most useful malingering techniques available: it's free. You can generate your own copy wherever you are; you need just a pencil and two pieces of paper. It takes a minute to administer. It is scored instantaneously. A positive score is generally meaningful and should lead to further investigation of why the performance is so poor. What are the pitfalls? A negative score generally means little, because the task demand is low (Faust & Auckley, 1998), because the test may not have face validity (thus, easily discerned to be an artifice), and because minimal coaching allows successful completion for the malingerer (Frederick et al., 1994). Positive scores may reflect significant neuropsychological impairment (e.g., see Schretlen et al., 1991). As noted by Vallabhajosula and van Gorp (2001), a successful malingering test "should be sensitive to faking but insensitive to genuine memory disorders," and the RMT does not appear to have fulfilled these criteria. Consequently, the RMT remains a useful procedure to investigate suspicious presentations, but it does not appear to have the capacity to be used reliably as a sole detection strategy for malingered cognitive impairment.

On the other hand, the WRT, and especially the process of comparing recall to recognition memory, has been demonstrated to be especially effective at identifying malingered cognitive impairment (Frederick, 2000; Greiffenstein et al., 1994, 1996; Haines & Norris, 2001). These studies typically modified Rey's approach by considering misrecognition errors in comparisons of recall and recognition memory. The DCT has not been shown to be an especially effective method of investigating malingered cognitive impairment, although Rey's strategy, performance curve analysis, has proven especially effective (Frederick, Crosby, & Wynkoop, 2000). It is likely the DCT could be improved with additional trials; six trials may be insufficient for reliable measurement of intention.

The caveats for the use of any procedure to evaluate malingered cognitive impairment are best summed up in the clinical wisdom of Rey:

> We see how it is possible to suspect the exaggeration and the simulation by using appropriately a battery of tests. It is sufficient to know the degree of intrinsic difficulty of each test well. However, the tests must be numerous in order to accumulate the suspicious signs: We will not be able to arrive at a conclusion on a single sign. Besides, it is always necessary to attribute these signs to the particular problem of the individual, to the traits of his personality, and to establish clearly what interests he may have in simulating disorders of a psychological nature. (Rey, 1958; p. 123)

## NOTE

Mme. Rey described to me many elaborate procedures Rey used for the assessment of feigning which have not been reported in American neuropsychological literature. Each essentially comprises a series of tasks which seem difficult but are not. Rey apparently plotted performance curves to demonstrate inconsistencies between performance accuracy and true difficulty levels.

## REFERENCES

Arnett, P. A., & Franzen, M. D. (1997). Performance of substance abusers with memory deficits on measures of malingering. *Archives of Clinical Neuropsychology, 12*, 513-518.

Arnett, P. A., Hammeke, T. A., & Schwartz, L. (1995). Quantitative and qualitative performance on Rey's 15-Item Test in neurological patients and dissimulators. *The Clinical Neuropsychologist, 9*, 17-26.

Association des amis d'André Rey (1967). *Hommage a André Rey*. Brussells: Dessart.

Back, C., Boone, K. B., Edwards, C., Parks, C., Burgoyne, K., & Silver, B. (1996). The performance of schizophrenics on three cognitive tests of malingering, Rey 15-Item Memory Test, Rey Dot Counting, and Hiscock Forced-Choice Method. *Assessment, 3*, 449-457.

Bernard, L. C., Houston, W., & Natoli, L. (1993). Malingering on neuropsychological memory tests: Potential objective indicators. *Journal of Clinical Psychology, 49*, 45-53.

Bernard, L. C., & Fowler, W. (1990). Assessing the validity of memory complaints: Performance on brain-damaged and normal individuals on Rey's task to detect malingering. *Journal of Clinical Psychology, 49*, 432-436.

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

DiCarlo, M. A., Gfeller, J. D., & Drury, J. A. (1996, October). Assessing feigned cognitive impairment with Rey's 15-Item Memory Test: Is this task easy or what? Poster presented at the Annual Meeting of the National Academy of Neuropsychology, New Orleans, LA.

Faust, D., & Auckley, M. A. (1998). Did you think it was going to be easy? Some methodological suggestions for the investigation and development of malingering detection techniques. In C. R. Reynolds (Ed.). *Detection of Malingering During Head Injury Litigation* (pp. 1-54). New York: Plenum.

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

Frederick, R. I. (2000a). A personal floor effect strategy to evaluate the validity of performance on memory tests. *Journal of Clinical and Experimental Neuropsychology, 22,* 720-730.

Frederick, R. I. (2000b). Mixed group validation: A method to address the limitations of criterion group validation in research on malingering detection. *Behavioral Sciences and the Law, 18,* 693-718.

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

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.

Goldberg, J. O., & Miller, H. R. (1986). Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints. *Journal of Clinical Psychology, 42,* 792-795.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures in a large clinical sample. *Psychological Assessment, 6,* 218-224.

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.

Griffin, G. A. E., Glassmire, D. M., Henderson, E. A., & McCann, C. (1997). Rey II: Redesigning the Rey Screening Test of Malingering. *Journal of Clinical Psychology, 53,* 757-766.

Griffin, G. A. E., Normington, J., & Glassmire, D. (1996). Qualitative dimensions in scoring the Rey Visual Memory Test of Malingering. *Psychological Assessment, 4,* 383-387.

Gudjonsson, G. H., & Shackleton, H. (1986). The pattern of scores on Raven's Matrices during "faking bad" and "non-faking" performance. *British Journal of Clinical Psychology, 25,* 35-41.

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.

Haines, M. E., & Norris, M. P. (2001). Comparing student and patient simulated malingerers' performance on standard neuropsychological measures to detect feigned cognitive deficits. *The Clinical Neuropsychologist, 15,* 171-182.

Hays, J. R., Emmons, J., & Lawson, K. A. (1993). Psychiatric norms for the Rey 15-Item Visual Memory Test. *Perceptual and Motor Skills, 76,* 1331-1334.

Iverson, G. L., & Franzen, M. D. (1996). Using multiple objective memory procedures to detect simulated malingering. *Journal of Clinical and Experimental Neuropsychology, 18,* 38-51.

Kaschel, R., Goldenberg, G., Goldstein, L. H., Risberg, J., Laaksonen, R., Gaillard, F., Perret, E., Tissot, R., & Oktem-Tanor, O. (1994). *Neuropsychology in Europe: A partial account of history and perspectives.* Paris: Approche Neuropsychologique des Apprentissages Chez L'Enfant.

Lambercier, M., & Rey, A. (1935). Contribution a l'etude de l'intelligence pratique chez l'enfant [Contribution to the study of practical intelligence in children]. *Archives de Psychologie, 25,* 1-59.

Lee, A., Boone, K. B., Lesser, I., Wohl, M., Wilkins, S., & Parks, C. (2000). Performance of older depressed patients on two cognitive malingering tests: False positive rates for the Rey 15-Item Memorization and Dot Counting Tests. *The Clinical Neuropsychologist, 14,* 303-308.

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

Lezak, M. D. (1983). *Neuropsychological assessment.* (2d ed.). New York: Oxford.

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

McKinzey, R. K., Podd, M. H., Krehbel, M. A., & Raven, J. (1999). Detection of malingering on Raven's Standard Progressive Matrices: A cross-validation. *British Journal of Clinical Psychology, 38,* 435-439.

Paul, D., Franzen, M. D., Cohen, S. H., & Fremouw, W. (1992). An investigation into the reliability and validity of two tests used in the detection of dissimulation. *International Journal of Clinical Neuropsychology, 14,* 1-9.

Rey, A. (1928). French philosophy in 1926 and 1927. *Philosophical Review, 37,* 527-556.

Rey, A. (1938). Le destruction progressive du cortex cérébral chez le rat et ses effets immédiats [The progressive destruction of the cerebral cortex in the rat and its immediate effects]. *Archives de Psychologie, 27,* 1-88.

Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique [The psychological examination in cases of traumatic encephalopathy]. *Archives de Psychologie, 28,* 286-340.

Rey, A. (1946). Epreuves de dessin: témoins du développement mental [Drawing tests as evidence of mental development]. *Archives de Psychologie, 31,* 369-380.

Rey, A. (1957). *Enfants Juifs Nord-Africains: Essai psycho-pédagogique a l'intention des educateurs [North-African Jewish children: Psycho-pedagogic essay for the use of educators].* Tel-Aviv: Departement pour l'Aliyah des Jeunes de l'Agence Juive.

Rey, A. (1958). *L'examen clinique en psychologie [The psychological examination].* Paris: Presses Universitaires de France.

Rogers, R., Harrell, E. H., & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review, 13,* 255-274.

Rose, F. E., Hall, S., & Szalda-Petree, A. D. (1998). A comparison of four tests of malingering and the effects of coaching. *Archives of Clinical Neuropsychology, 13,* 349-363.

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.

Simon, M. J. (1994). The use of the Rey Memory Test to assess malingering in criminal defendants. *Journal of Clinical Psychology, 50,* 667-672.

Taylor, E. M. (1959). *The appraisal of children with cerebral deficits.* Cambridge, MA: Harvard University Press.

Vallabhajosula, B., & van Gorp, W. G. (2001). Post-*Daubert* admissibility of scientific evidence on malingering of cognitive deficits. *Journal of American Academy of Psychiatry and Law, 29,* 207-215.

Wintsch, J. (1935). A propos d'une etude sur l'intelligence pratique chez l'enfant [A study of practical intelligence in children]. *Zeitschrift fuer Kinderpsychiatrie, 2,* 146-149.