JOURNAL OF RESEARCH IN PERSONALITY 13, 245–265 (1979)

A General Scheme for the Analytic Decomposition of Objective Test Scores: Illustrative Demonstrations Using the Rod-and-Frame Test and the Müller–Lyer Illusion

LEWIS R. GOLDBERG

University of Oregon and Institute for the Measurement of Personality

Objective tests of personality typically include a number of items or trials; the total score on the test is the sum of the subject's "correct" responses across all such trials. Normally, the trials are varied systematically across various facets of the test design, so that the total score represents a composite measure of accuracy averaged across these test facets. However, since only one score is computed for each subject, some potentially important kinds of individual differences—namely all those associated with each particular variation in the test design—are treated solely as measurement unreliability. Such a psychometric stance may serve to obscure more differentiated types of individual differences, with the result that composite scores from trials based on one type of experimental design may not be highly related to such scores from trials using a somewhat different design. The present paper presents a general procedure for scoring objective tests more analytically. To illustrate this general rationale, and to demonstrate its potential utility, data have been reanalyzed from two previous studies, one using the Rod-and-Frame test, the other the Müller–Lyer illusion. In both cases, the traditional global accuracy score did not correlate significantly with other theoretically related variables, while a number of component scores were quite highly related.

While early attempts to develop "objective" or "maximum performance" measures of personality date back to the early 1920s (see Burtt, 1923; Travis, 1925), the past decade has witnessed a burgeoning interest in the development of objective personality tests. These efforts have been stimulated by the apparent fruitfulness of research on perceptual and cognitive styles (e.g., Blake & Ramsey, 1951; Klein, 1970; Witkin, Lewis,

The author is deeply indebted to Joop Hettema for initially posing this problem, for his help in formulating a first conceptual scheme, and for providing the RFT data here reanalyzed. The author is also indebted to Mucella Ormanli for providing the MLI data, and to Willem K. B. Hofstee. Thom Bezembinder, Jerry S. Wiggins, and Donald W. Fiske for their many cogent suggestions for improving the manuscript. This project was supported by Grants MH 12972 and MH 30325 from the National Institute of Mental Health, U.S. Public Health Service. Requests for reprints should be sent to Lewis R. Goldberg, Institute for the Measurement of Personality, 1201 Oak St., Eugene, OR 97401.

0092-6566/79/020245–21$02.00/0

Hertzman, Machover, Meissner, & Wapner, 1954; Witkin, Dyk, Faterson, Goodenough, & Karp, 1962), by the publication of objective test compendia by Raymond B. Cattell and his collaborators (e.g., Cattell & Warburton, 1967; Hundleby, Pawlik, & Cattell, 1965), and especially by the recent wave of pessimism concerning the utility of self-report scales and inventories (see Mischel, 1968; Wallace, 1966).

The term "objective," as here employed, includes two critical characteristics; tests are objective to the extent that (a) their scoring can be automated (that is, to the extent to which there is perfect consensus between diverse scorers), and (b) testees can deliberately distort their true scores in only one direction (that is, to the extent to which testees can fake "bad" but not "good"). The MMPI is objective by the first standard but not by the second. The Bender–Gestalt Test is objective by the second standard, but not by the first. The Scholastic Aptitude Test and the Graduate Record Examination are objective by both standards, as are such measures of cognitive style as the Rod-and-Frame and the Embedded Figures tests.

Objective tests of personality, like those of aptitude and achievement, typically include a number of questions, items, or trials; the total score on the test is the sum (or average) of the subject's "correct" responses across all such trials. Normally the trials are varied systematically across various facets of the test design, so that the total score represents a composite measure of accuracy averaged across these test facets. Yet, only one score is computed for each subject, regardless of the number and nature of the facets built into the test design. As a consequence, some potentially important kinds of individual differences—namely all those associated with each particular variation in the test design—are treated solely as intraindividual variability or measurement unreliability.

While such a psychometric stance could serve to increase the across-facet generality of the composite measure, it could also serve to obscure more differentiated types of individual differences, with the result that composite scores from trials based on one type of experimental design may not be highly related to such scores from trials using a somewhat different design. By scoring the more-differentiated component scores *in addition to* the composite measure, it should be possible to align the results from diverse investigations in a more coherent fashion. This argument is essentially analogous to that made some time ago by Cronbach, when he noted that the global measures of judgmental accuracy used in studies of person perception were aggregate indexes that could, and should, be decomposed into more meaningful component scores (see Cronbach, 1955; Gage & Cronbach, 1955).

To illustrate this general rationale, and to demonstrate its potential utility, we will reanalyze the data from two previous studies, one using the Rod-and-Frame Test, the other the Müller–Lyer Illusion.

THE ROD-AND-FRAME TEST

The prototype of objective personality measures is the Rod-and-Frame Test (RFT), originally reported by Witkin and Asch as a by-product of some basic experimental studies of spatial orientation, especially perception of the upright (Asch & Witkin, 1948a,b; Witkin, 1949a,b; Witkin & Asch, 1948a,b). In these early experiments, Witkin and Asch observed that while the perception of the upright was generally determined both by internal (proprioceptive) and by external (visual) cues, there were large individual differences among subjects in their reliance on external cues when these visual cues were experimentally placed in opposition to the internal ones. For example, when seated in a dark room and asked to vertically position a luminous rod within a luminous square frame which was tilted to one side, subjects generally positioned the rod in some compromise position between the true vertical and the "visual vertical" provided by the tilted frame. However, some subjects appeared to disregard the cues provided by the tilted frame, while others appeared to rely almost completely on such visual cues. Individual differences in accuracy of rod positioning from the true vertical, which were found to relate to accuracy in a variety of similar spatial tasks, were originally conceptualized as a trait of "field dependence" (Witkin et al., 1954). Later, after further investigations suggested that RFT accuracy scores were also related to analogous measures from a number of other tasks (especially the Embedded Figures Test), the concept was broadened and retitled "psychological differentiation" (Witkin et al., 1962).

Perhaps because of their initial focus on perception of the upright, Witkin and Asch—as well as most of the investigators who later used the RFT technique—elected to compute only a single global accuracy score for each subject, namely the sum of the absolute values of the deviations of the subject's rod settings from the true vertical across all of the RFT trials. However, such an aggregate index necessarily obscures such factors in RFT performance as the direction of the deviations, their sequential order within the set of RFT trials, and their relationship to the initial rod and frame settings. A finer-grained analysis of the potentially complex processes involved in the RFT may serve to clarify some of the controversial issues regarding its construct validity (e.g., Gruen, 1957; Vernon, 1972; Zigler, 1963), especially the relationship between individual differences in RFT performance and general intelligence, the perceptual closure factors, and responses to optical illusions.

For despite a vigorous plea for the standardization of performance tests of personality (Krathwohl & Cronbach, 1956), variations in RFT apparatus and procedures are legion. Uncontrolled variations between different RFT studies include: (a) the size of the square frame; (b) the distance between the subject and the frame; (c) the occasional use of a head clamp to restrict subjects' movements; (d) the occasional use of goggles to eliminate peripheral vision; (e) the occasional elimination of a chair (movable or not) in favor of standing erect (sometimes on a tiltable board); (f) the use of instructions to set the rod at the vertical, the horizontal, or both; (g) the initial positions of the frame; (h) the initial positions of the rod; and (i) the number of trials run under each experimental condition.

However, with very few exceptions (e.g., Benfari & Vitale, 1965; Cabe, 1968a; Coan, 1964; Hettema, 1968; Pressey, 1967), the single most standard aspect of the RFT has been its scoring, perhaps because RFT investigators have been influenced by the following argument:

> There are a number of possible alternative scoring methods for the RFT and for the other orientation tests. For example, the mean error in the direction of the tilt of the frame has been used by some workers instead of the absolute error. These alternatives may be useful for specific research problems, but we have not found a scoring method that has greater construct validity as a measure of field dependence than does the absolute error (Witkin et al., 1962, p. 37, footnote 2).

Such an argument would seem more compelling if the evidence for the assertion were somewhere presented. Until such evidence has accrued, it does not seem unreasonable to propose that RFT investigators analyze a set of component scores, along with the global index.

PARAMETERS OF RFT PERFORMANCE AND SCORING

The traditional RFT global accuracy score represents an aggregate pattern of performance across such situational and instructional factors as (a) the subject's body orientation (tilted vs erect), (b) the initial frame position (tilted to the right vs to the left), (c) the initial rod position (tilted to the right vs to the left), and (d) the sequential order within the total set of trials (early trials vs late trials). In addition, the index might be averaged across such factors as (e) the subject's physical stance (seated vs standing), and (f) instructions to position the rod to the vertical vs the horizontal. Each of these factors can be viewed as exerting a "main effect" on individual differences in RFT performance, as can the various two-way, three-way, and higher-order interactions among these factors (see Table 1). Among these latter effects, the body orientation by frame position interaction (e.g., subject and frame both tilted to the same side vs subject and frame differing in orientation) and the analogous frame position by rod position interaction effect would a priori seem to be worthy of particular scrutiny.

As indicated in Table 1, at least six types of scores can be computed for each of these factors (and each of the interactions). First of all, *accuracy* (*A*) scores can be differentiated from measures of *bias* (*B*). All accuracy scores are averages of the *absolute* values of the discrepancies between the final rod setting and some standard (e.g., the true vertical); measures of bias are analogous averages of the *algebraic* (i.e., signed) discrepan-

TABLE 1
A GENERAL DESIGN FOR THE ANALYSIS OF RFT PERFORMANCE

	Accuracy (*A*)					
	True vertical (*T*)		Personal vertical (*P*)		Bias (*B*)	
Main effects	Directional (*D*)	Nondirectional (*N*)	Directional (*D*)	Nondirectional (*N*)	Directional (*D*)	Nondirectional (*N*)
Body orientation (*O*)	*OATD*	*OATN*	*OAPD*	*OAPN*	*OBD*	*OBN*
Frame position (*F*)	*FATD*[a]	*FATN*[a]	*FAPD*[a]	*FAPN*[a]	*FBD*[a]	*FBN*[a]
Rod position (*R*)	*RATD*	*RATN*	*RAPD*	*RAPN*	*RBD*	*RBN*
Sequence (*S*)	*SATD*[a]	*SATN*[a]	*SAPD*[a]	*SAPN*[a]	*SBD*[a]	*SBN*[a]
Interaction effects						
$O \times F$ (*G*)	*GATD*	*GATN*	*GAPD*	*GAPN*	*GBD*	*GBN*
$O \times R$ (*H*)	*HATD*	*HATN*	*HAPD*	*HAPN*	*HBD*	*HBN*
$O \times S$ (*I*)	*IATD*	*IATN*	*IAPD*	*IAPN*	*IBD*	*IBN*
$F \times R$ (*J*)	*JATD*	*JATN*	*JAPD*	*JAPN*	*JBD*	*JBN*
$F \times S$ (*K*)	*KATD*[a]	*KATN*[a]	*KAPD*[a]	*KAPN*[a]	*KBD*[a]	*KBN*[a]
$R \times S$ (*L*)	*LATD*	*LATN*	*LAPD*	*LAPN*	*LBD*	*LBN*
$O \times F \times R$ (*U*)	*UATD*	*UATN*	*UAPD*	*UAPN*	*UBD*	*UBN*
$O \times F \times S$ (*V*)	*VATD*	*VATN*	*VAPD*	*VAPN*	*VBD*	*VBN*
$O \times R \times S$ (*W*)	*WATD*	*WATN*	*WAPD*	*WAPN*	*WBD*	*WBN*
$F \times R \times S$ (*X*)	*XATD*	*XATN*	*XAPD*	*XAPN*	*XBD*	*XBN*
$O \times F \times R \times S$ (*Y*)	*YATD*	*YATN*	*YAPD*	*YAPN*	*YBD*	*YBN*
Total effect (*Z*)	*ZAT*[a]		*ZAP*[a]		*ZBD*[a]	*ZBN*[a]

[a] Indexes included in the empirical study.

cies. Moreover, among accuracy scores, deviations can be computed from the "true" (T) vertical or from the "personal" (P) vertical (the algebraic mean of the individual's final rod settings over all trials). And, finally, most of these indices can be calculated either as directional (D) or as nondirectional (N) scores. Directional scores retain their algebraic sign, while each nondirectional score is the absolute value of its directional analogue.

A SIMPLE COMPUTATIONAL EXAMPLE

To make these distinctions clear, let us now describe how these scores may be calculated, using a simplified example. Assume that there are only four RFT trials and that the initial frame settings are alternated from left to right between each trial; thus, the frame is initially set to the right on trials 1 and 3 and to the left on trials 2 and 4. Consider the first two trials as the "earlier" trials, and the last two as the "later" trials. The initial data consist of four *signed* discrepancies between the final rod setting for each trial and the true vertical; arbitrarily, assume that errors toward the subject's right are given positive signs, while those toward his/her left are given negative signs.

Let us designate the four responses as C_1, C_2, C_3, and C_4. The "personal" vertical ($\bar{C}$) is therefore defined as:

$$\bar{C} = \tfrac{1}{4} \sum_{i=l}^{4} C_i. \tag{1}$$

Discrepancies from the personal vertical, designated E_i, are defined as:

$$E_i = C_i - \bar{C}. \tag{2}$$

Finally, let us denote the absolute values of C_i as M_i, and analogously denote the absolute values of E_i as Q_i:

$$M_i = |C_i|, \tag{3}$$

$$Q_i = |E_i|. \tag{4}$$

The computing formulas for each of the six "frame position" (F) effects are given in Eqs. (5)–(10).

$$FATD = M_1 - M_3 + M_2 - M_4, \tag{5}$$

$$FATN = |M_1 - M_3 + M_2 - M_4|, \tag{6}$$

$$FAPD = Q_1 - Q_3 + Q_2 - Q_4, \tag{7}$$

$$FAPN = |Q_1 - Q_3 + Q_2 - Q_4|, \tag{8}$$

$$FBD = C_1 - C_3 + C_2 - C_4, \tag{9}$$

$$FBN = |C_1 - C_3 + C_2 - C_4|. \tag{10}$$

The analogous computing formulas for each of the six "sequential" (S) effects are given in Eqs. (11)–(16).

$$SATD = M_1 + M_2 - M_3 - M_4, \quad (11)$$
$$SATN = |M_1 + M_2 - M_3 - M_4|, \quad (12)$$
$$SAPD = Q_1 + Q_2 - Q_3 - Q_4, \quad (13)$$
$$SAPN = |Q_1 + Q_2 - Q_3 - Q_4|, \quad (14)$$
$$SBD = C_1 + C_2 - C_3 - C_4, \quad (15)$$
$$SBN = |C_1 + C_2 - C_3 - C_4|. \quad (16)$$

The computing formulas for each of the six "interaction" (K) effects are given in Eqs. (17)–(22).

$$KATD = M_1 + M_4 - M_2 - M_3, \quad (17)$$
$$KATN = |M_1 + M_4 - M_2 - M_3|, \quad (18)$$
$$KAPD = Q_1 + Q_4 - Q_2 - Q_3, \quad (19)$$
$$KAPN = |Q_1 + Q_4 - Q_2 - Q_3|, \quad (20)$$
$$KBD = C_1 + C_4 - C_2 - C_3, \quad (21)$$
$$KBN = |C_1 + C_4 - C_2 - C_3|. \quad (22)$$

Finally, the computing formulas for each of the four "total" (Z) effects are given in Eqs. (23)–(26).

$$ZAT = M_1 + M_2 + M_3 + M_4, \quad (23)$$
$$ZAP = Q_1 + Q_2 + Q_3 + Q_4, \quad (24)$$
$$ZBD = C_1 + C_2 + C_3 + C_4, \quad (25)$$
$$ZBN = |C_1 + C_2 + C_3 + C_4|. \quad (26)$$

CORRELATES OF RFT COMPONENT SCORES: AN ILLUSTRATIVE STUDY

The use of the RFT in personality assessment is predicated on the assumption that RFT scores relate to a host of other cognitive, personality and/or perceptual measures. Consequently, to the extent that the proposed RFT component scores are useful additions to the global accuracy index (ZAT), these scores should correlate more highly—and/or more consistently—with such non-RFT indexes. As a preliminary test of this hypothesis, the data from a study by Hettema (1968) were reanalyzed, using 22 indexes from the new scoring design.

Subjects. The subjects were 70 young male adults from the Dutch Army, who are further described in "Study I" of Hettema (1968).

RFT procedures and scores. The RFT was administered under the experimenter-operated mode. Subjects were seated erect 3 m from the RFT apparatus. Twelve trials were administered, with the frame alternately rotated 28° to the right and 28° to the left of the true vertical. The rod was always set 28° from the true vertical, to the opposite side of the tilt of the frame. Algebraic (signed) deviations of the rod from the true vertical were recorded at the end of each of the 12 trials.

From these initial 12 deviation values, the 22 scores which are marked by an asterisk in Table 1 were computed. Of these 22 indexes, 6 were based upon the initial frame position (the six even-number trials, in which the frame was initially tilted to the right vs the six odd-numbered trials, in which it was tilted to the left): *FATD, FATN, FAPD, FAPN, FBD, FBN.* (Since the initial rod and frame positions were completely confounded in this study, it was not possible to include any independent scores based on the initial rod settings; consequently, these six "frame position effects" are actually "rod-by-frame interaction effects" in the more complex design displayed in Table 1. However, for conceptual simplicity, they will be considered as if they were solely "frame effects" in the present analyses.) Six scores based upon "sequential effects" (the first six trials vs the last six trials) were included: *SATD, SATN, SAPD, SAPN, SBD,* and *SBN*. Six "frame position by sequence" interaction effects (trials, 1, 3, 5, 8, 10, and 12 vs trials 2, 4, 6, 7, 9, and 11) were included: *KATD, KATN, KAPD, KAPN, KBD,* and *KBN*. And, finally, the four "total" RFT scores were included: *ZAT, ZAP, ZBD,* and *ZBN*.

Other measures. Ten additional cognitive and perceptual tests were included in the study. Eight of these tests, which were borrowed from Thurstone's (1944) factorial analysis of perceptual tasks, were included as marker variables for four perceptual factors: *Factor I. Speed and Strength of Closure* (Street Gestalt Test, Gottschaldt Hidden Figures, PMA—Spatial Test); *Factor II. Flexibility of Closure* [Hidden Pictures Test, PMA—Reasoning Test (Letter Series)]; *Factor III. Optical Illusions* (Müller—Lyer Illusion, Poggendorff Illusion); and *Factor IV. Perceptual Speed* (Peripheral Span—Test B). Seven of these tests were adapted for group administration, while the Peripheral Span Test was administered individually. All test materials were copied from Thurstone (1944), except for the Hidden Pictures Test for which new designs were constructed. In addition, a measure of the cognitive style labeled "leveling vs sharpening" (the Squares Test) and a measure of general verbal aptitude (a Verbal Analogies Test) were included in the study; for a description of these two measures, see Goldberg (1972). All tests were scored so that high scores reflected less accurate perception (i.e., more errors).

Analyses. While both linear (Pearson r) and nonlinear (η) correlations were computed among these 32 measures (the 22 RFT scores and the 10 additional indexes), only the linear relationships will be reported here. For the complete findings, see Goldberg (1972).

Results and Discussion

The means, standard deviations, and intercorrelations among the 22 RFT scores are presented in Goldberg (1972), along with the intercorrelations among the 10 additional cognitive and perceptual tests. The most substantial relationships among the latter measures, which could well be attributed to general scholastic aptitude, were found among the Verbal Analogies, PMA Reasoning, Gottschaldt, Hidden Figures, and PMA Spat-

TABLE 2

CORRELATIONS BETWEEN THE 22 RFT SCORES AND THE 10 ADDITIONAL TESTS (N = 70)

	ZBD	*ZBN*	*ZAT*	*ZAP*	*FBD*	*FBN*	*FATD*	*FATN*	*FAPD*	*FAPN*
Square Test	−.04	−.11	−.04	.03	−.01	.03	.08	.06	.04	−.02
Peripheral Span	−.15	−.22	−.16	−.09	−.07	−.09	−.09	−.19	.03	−.02
Street Gestalt	−.10	−.15	.11	.26*	.22	.30**	.15	.29*	.10	−.04
Hidden Pictures	−.13	−.24*	.08	.21	.10	.22	.03	.06	−.15	.08
Poggendorff	.02	.07	.06	.06	.32**	.05	.27*	.13	.04	.03
Müller–Lyer	.12	−.01	−.11	−.12	−.30**	−.13	.08	−.10	.01	−.08
PMA Spatial	−.06	−.03	.11	.19	.09	.19	.00	.22	.03	.01
Gottschaldt	−.08	−.20	.10	.21	−.09	.23*	.03	.11	.01	.08
PMA Reasoning	−.19	−.29*	−.02	.11	−.06	.12	.07	−.06	.02	.01
Verbal Analogies	−.10	−.20	.14	.22	−.24*	.23*	−.09	−.02	.10	.02

* $p < .05$.
** $p < .01$.

ial Tests. In addition, the PMA Spatial Test was significantly correlated with the Hidden Pictures Test and the Poggendorff Illusion; the Müller–Lyer Illusion was significantly related to the PMA Reasoning Test; and the Street Gestalt Test was significantly related to the Hidden Pictures Test. However, three-quarters of the intercorrelations among these 10 tests were not significantly greater than zero, including all of the intercorrelations among the putative marker tests for each of the two closure factors.

Table 2 presents the correlations between the 22 RFT scores and the 10 additional cognitive and perceptual tests. Over 10% of the correlations between these two sets of measures were significantly greater than zero ($p < .05$). The highest of these relationships was that between one of the RFT nondirectional accuracy scores based upon sequential effects (*SAPN*) and the Gottschaldt Hidden Figures Test ($r = .35$). This same RFT index also had significant correlations with the PMA Reasoning, Verbal Analogies, and Street Gestalt tests. At the other extreme, seven RFT scores failed to correlate significantly with *any* of the additional measures: *FAPD*, *FAPN*, *SBN*, *KATN*, *KAPN*, *ZBD*, and—of most importance—*ZAT*, the traditional RFT global accuracy score.

Thus, an investigator who computed only the traditional RFT global accuracy score (*ZAT*) using this same sample of subjects and this same battery of perceptual and cognitive tests would have been forced to conclude that there was *no* relationship between RFT performance and scores on the other ten tasks. Consequently, the obtained pattern of cross-task linkages displayed in Table 2 provides some preliminary evidence for the usefulness of these RFT component scores. However, the use of 22 scores, instead of one global index, greatly increases the probability that some of these significant relationships arose from chance vagaries of the particular sample under study. Only future replications of these procedures, ideally based upon larger samples of subjects and

TABLE 2 (continued)

SBD	*SBN*	*SATD*	*SATN*	*SAPD*	*SAPN*	*KBD*	*KBN*	*KATD*	*KATN*	*KAPD*	*KAPN*
−.05	.18	.21	.10	−.08	−.01	−.01	−.15	.12	.07	.05	.06
.04	−.02	.24*	.08	.14	.04	−.07	.11	−.03	−.03	−.04	.06
.06	.12	.22	.16	−.19	.26*	−.28*	.16	.04	.13	.12	.13
.12	.13	.03	.11	.02	.19	.00	.10	.15	.03	.15	.17
.01	.06	−.10	.08	−.13	.10	−.10	.22	−.00	.12	.08	.08
−.11	−.16	.25*	−.14	.26*	−.10	.04	−.03	−.29*	−.10	−.25*	−.21
.02	.15	−.10	.21	−.09	.19	.07	.16	.11	.21	.20	.19
−.18	−.05	−.14	.26*	−.05	.35**	.16	.24*	.05	.00	−.04	.01
−.11	−.09	−.16	.21	−.13	.28*	−.21	.31**	.02	.05	.00	−.00
−.23*	−.18	−.15	.20	−.06	.26*	−.17	.18	−.06	−.00	−.13	−.02

utilizing more elements from the experimental design displayed in Table 1, will demonstrate whether the use of component scores improves our theoretical understanding of these individual differences.

Moreover, if this analytic scheme proves useful for the RFT, it should be equally fruitful when applied to *other* objective tests—including those used in this illustrative study. Indeed, the most useful type of across-test correlational pattern would be based on the same, or similar, component scores from *each* of the objective tests under study. Unfortunately, the data now available from Hettema's (1968) study are not complete enough to permit parallel analyses of these other objective tests. Seemingly, however, most of these tests—and especially the perceptual illusions—should lend themselves rather easily to similar analyses. As an illustration, data from a study of the Müller–Lyer Illusion have been reanalyzed.

A SECOND EXAMPLE: THE MÜLLER–LYER ILLUSION

While the Rod-and-Frame Test is of relatively recent vintage, scientific reports on the Müller–Lyer Illusion (MLI) date back well into the 19th century; for an extensive review of this early work, as well as of more current studies, see Robinson (1972). Moreover, while the RFT has been employed primarily as an instrument for measuring individual differences, the MLI has only rarely been used for this purpose, in spite of repeated observations of large individual differences in the initial strength of the illusion and of large differential effects of prolonged practice (see Eysenck & Slater, 1958; Gardner, 1961). In a recent study, the MLI was included along with the Bender–Gestalt and the Trail Making tests as potential predictors of psychosis in 8- to 12-year-old children (Ormanli, 1975).

Subjects. The subjects for this study were 60 children, half of whom had been diagnosed as psychotic and half of whom served as normal controls. The normal sample consisted of 15 boys and 15 girls. Teachers selected the initial candidates for testing by excluding students with either very high or very low intelligence scores. The psychotic sample was composed of 22 boys and 8 girls, approximately equally divided between two hospitals, and matched in age with subjects in the control sample. The great majority of the subjects were Caucasians; there were two black children in the normal sample and three in the psychotic sample. No patients were included in the psychotic sample if they were acutely ill, disoriented, confused, or uncooperative at the time of testing; they were tested either before or at the very beginning of any treatment. The diagnosis for each of these patients was based on the information in their hospital records, including a detailed medical history, a psychiatric examination, a psychological evaluation, and any neurological reports.

MLI testing procedures. All subjects were tested individually, using six different administration orders for the three tests employed. Since the order of administration had no significant effect on test performance, it will not be further discussed.

The MLI apparatus consisted of two sections, each made of white cardboard 213 mm in length. On one of these sections, the Standard, there was a 100-mm shaft, at the ends of which were outpointing fins 20 mm in length, each set at a 45° angle to the shaft. On the other section, the Comparison, there was a shaft that was adjustable in length (up to a maximum of 200 mm), with inpointing fins at one end. All lines were drawn in black ink, 1 mm wide. The Comparison section could easily slide into the Standard section by pulling it out or pushing it in. A ruler was affixed to the reverse side of the Comparison section, allowing readings to be taken in millimeters.

Before the MLI was administered, an easy task was given in order to make the child feel confident and relaxed. Then each subject was given 20 trials on the MLI, 10 trials in which the Comparison stimulus was initially set 100 mm longer than the Standard, and 10 trials in which it was initially set 100 mm shorter. On 5 of each of these 10 trials, the Standard was to the subjects' right, and on the other 5 to their left. Thus, there were four test conditions: (A) Standard on the right, and Comparison longer (five trials); (B) Standard on the right, and Comparison shorter (five trials); (C) Standard on the left, and Comparison longer (five trials); and (D) Standard on the left, and Comparison shorter (five trials). The four test conditions were administered in the following order: D A B C B C B A D C B A D C B A D A D C. Subjects were instructed to make the Comparison shaft equal in length to the Standard shaft, by adjusting the Comparison section.

Scoring the MLI. MLI scores were computed from two experimental designs, the original based on 20 trials, and a simpler and better balanced one based on 16 trials. Both designs are presented in Table 3. Since the results from the two were quite similar, only the analyses from the 16-trial design will be reported here.

As Table 3 indicates, there are three major factors that might affect test responses in this experimental design: (a) the *Orientation* of the Standard (on the subject's left vs right), (b) the *Length* of the Comparison shaft in its initial setting (shorter vs longer than the Standard), and (c) the *Sequence* in the trials (the early vs the late trials). Table 4 lists the 46 scores that are derivable from this experimental design, when each of the three main factors and their

TABLE 3

THE EXPERIMENTAL DESIGN EMPLOYED IN THE MLI STUDY: THE ORIGINAL 20 TRIALS AND THE 16-TRIAL SUBSET

Original Trial No.	Condition	Orientation of standard	Length comparison initially set	Sequence		16-trial No.
				20 trials	16 trials	
1	D	Left	Shorter	Early	Early	1
2	A	Right	Longer	Early	Early	2
3	B	Right	Shorter	Early	Early	3
4	C	Left	Longer	Early	Early	4
5	B	Right	Shorter	Early	Early	5
6	C	Left	Longer	Early	Early	6
7	B	Right	Shorter	Early	—	—
8	A	Right	Longer	Early	Early	7
9	D	Left	Shorter	Early	Early	8
10	C	Left	Longer	Early	Late	9
11	B	Right	Shorter	Late	Late	10
12	A	Right	Longer	Late	Late	11
13	D	Left	Shorter	Late	Late	12
14	C	Left	Longer	Late	Late	13
15	B	Right	Shorter	Late	Late	14
16	A	Right	Longer	Late	Late	15
17	D	Left	Shorter	Late	Late	16
18	A	Right	Longer	Late	—	—
19	D	Left	Shorter	Late	—	—
20	C	Left	Longer	Late	—	—

interactions are crossed with the six types of measures that can be derived from MLI performance [directional and nondirectional measures of bias, accuracy from the true standard (in this case, 100 mm), and accuracy from the subject's personal standard].

To compute these 46 scores for each subject, his/her responses to each of the 16 trials, here designated X_i ($i = 1, \ldots, 16$), were first used to generate three transformed vectors:

(1) *Bias* (*C*): $C_i = 100 - X_i$ (27)

(2) *Accuracy from the True Standard* (*M*): $M_i = |C_i|$ (28)

The "Personal Standard" $\bar{C} = 1/16 \sum_{i=1}^{16} C_i$ (29)

The Discrepancy from the Personal Standard $E_i = C_i - \bar{C}$ (30)

(3) *Accuracy from the Personal Standard* (*Q*): $Q_i = |E_i|$ (31)

To generate the six *Orientation* (*O*) scores, the sum of the subject's responses to trials on which the Standard was on the left (b) were subtracted from the sum of those on the right (a):

$$OATD = M_a - M_b, \quad (32)$$
$$OATN = |M_a - M_b|, \quad (33)$$
$$OAPD = Q_a - Q_b, \quad (34)$$
$$OAPN = |Q_a - Q_b|, \quad (35)$$
$$OBD = C_a - C_b, \quad (36)$$
$$OBN = |C_a - C_b|. \quad (37)$$

To generate the six *Length* of Comparison (*L*) scores, the sum of the subject's responses to trials on which the Comparison was initially set longer than the Standard (d) were subtracted from the sum of those on which it was set shorter (c):

$$LATD = M_c - M_d, \quad (38)$$
$$LATN = |M_c - M_d|, \quad (39)$$
$$LAPD = Q_c - Q_d, \quad (40)$$
$$LAPN = |Q_c - Q_d|, \quad (41)$$
$$LBD = C_c - C_d, \quad (42)$$
$$LBN = |C_c - C_d|. \quad (43)$$

To generate the six *Sequence* (*S*) scores, the sum of the subject's responses to the late trials (f) were subtracted from the sum of those to the early trials (e):

$$SATD = M_e - M_f, \quad (44)$$
$$SATN = |M_e - M_f|, \quad (45)$$
$$SAPD = Q_e - Q_f, \quad (46)$$
$$SAPN = |Q_e - Q_f|, \quad (47)$$
$$SBD = C_e - C_f, \quad (48)$$
$$SBN = |C_e - C_f|. \quad (49)$$

Analogously, each of the six $O \times L$ *interaction* (*G*) scores were generated by taking those trials on which the Orientation was to the left and the comparison was shorter, plus those trials on which the Orientation was to the right and the Comparison was longer (i.e., trials 1, 2, 7, 8, 11, 12, 15, 16 in the 16-trial design displayed in Table 3) and contrasting them with those trials on which the Orientation was to the left and the Comparison was longer, plus those trials on which the Orientation was to the right and the Comparison was shorter (i.e., trials 3, 4, 5, 6, 9, 10, 13, 14). To generate each of the six $O \times S$ *interaction* (*H*) scores, early trials with Orientation to the left, plus late trials with Orientation to the right (i.e., trials 1, 4, 6, 8, 10, 11, 14, 15) were contrasted with early trials with Orientation to the right, plus late trials with Orientation to the left (i.e., trials 2, 3, 5, 7, 9, 12, 13, 16). Each of the six $L \times S$ *interaction* (*I*) scores were generated by contrasting early trials on which the Comparison

TABLE 4
A GENERAL DESIGN FOR THE ANALYSIS OF PERFORMANCE ON THE MÜLLER–LYER ILLUSION

	Accuracy (*A*)					
	True standard (*T*)		Personal standard (*P*)		Bias (*B*)	
	Directional (*D*)	Nondirectional (*N*)	Directional (*D*)	Nondirectional (*N*)	Directional (*D*)	Nondirectional (*N*)
Main effects						
Orientation of standard (*O*)	*OATD*	*OATN*	*OAPD*	*OAPN*	*OBD*	*OBN*
Length of comparison (*L*)	*LATD*	*LATN*	*LAPD*	*LAPN*	*LBD*	*LBN*
Sequence (*S*)	*SATD*	*SATN*	*SAPD*	*SAPN*	*SBD*	*SBN*
Interaction effects						
$O \times L$ (*G*)	*GATD*	*GATN*	*GAPD*	*GAPN*	*GBD*	*GBN*
$O \times S$ (*H*)	*HATD*	*HATN*	*HAPD*	*HAPN*	*HBD*	*HBN*
$L \times S$ (*I*)	*IATD*	*IATN*	*IAPD*	*IAPN*	*IBD*	*IBN*
$O \times L \times S$ (*J*)	*JATD*	*JATN*	*JAPD*	*JAPN*	*JBD*	*JBN*
Total effect (*Z*)	*ZAT*		*ZAP*		*ZBD*	*ZBN*

was shorter, plus late trials on which it was longer (i.e., trials 1, 3, 5, 8, 9, 11, 13, 15) with early trials on which it was longer plus late trials on which it was shorter (i.e., trials 2, 4, 6, 7, 10, 12, 14, 16). To generate the $O \times L \times S$ *interaction* (J) scores, trials 1, 2, 7, 8, 9, 10, 13, 14 were contrasted with trials 3, 4, 5, 6, 11, 12, 15, 16.

Finally, the four *total* (Z) scores were generated:

$$ZAT = \sum_{i=1}^{16} M_i, \tag{50}$$

$$ZAP = \sum_{i=1}^{16} Q_i, \tag{51}$$

$$ZBD = \sum_{i=1}^{16} C_i, \tag{52}$$

$$ZBN = |ZBD|. \tag{53}$$

Other measures: The Bender–Gestalt and Trail Making Tests. The Bender–Gestalt Test (BGT) consists of nine cards, on each of which is printed a simple figure, which the subject is instructed to copy. In this study, each design was copied on a separate sheet of paper, and the time required to copy each design was recorded by the experimenter. The Trail Making Test (TMT) has two parts, each including a short sample followed by the actual test. *Part A* consists of a page on which 15 circles are randomly distributed; in each of the circles is a number from 1 to 15. The subject is instructed to begin with the circle numbered 1 and to connect the circles by pencil in their correct numerical order. *Part B* consists of a similar page of circles, eight numbered from 1 to 8 and seven labeled from A to G. The subject is instructed again to connect the circles, proceeding this time in the order: 1, A, 2, B, . . . , 15. As with the BGT, the time required to complete each part of the TMT was recorded by the experimenter.

Results and Discussion

Since the Standard was 100 mm in length, the illusion should tend to produce Comparison settings that were *less* than 100 mm. Indeed, 100% of the responses from the normal sample and 95% of those from the psychotic sample were less than 100 mm. Table 5 presents the means and standard deviations, in millimeters, for each of the 46 MLI scores. The means for the total scores (which are summed across all 16 trials) when divided by 16 indicate the strength of the illusion for the average subject on the average trial; such values for the total accuracy (*ZAT*) and bias (*ZBD* and *ZBN*) scores were approximately 23 mm, indicating a quite substantial general illusion effect—the average subject setting the Comparison shaft nearly 25% shorter than the Standard.

The intercorrelations among the 46 MLI scores are available from the author. Since in this sample every subject showed some effect of the illusion (that is, the average response of every subject was below 100 mm), the directional and nondirectional total bias scores (*ZBD* and *ZBN*) were here identical. Moreover, because of this strong source of common bias, the correlations between most of the directional bias scores and the corresponding directional measures of accuracy from the true standard were close to unity (e.g., the correlation between *ZAT* and *ZBD* was .99;

TABLE 5
MEANS AND STANDARD DEVIATIONS FOR EACH OF THE 46 MÜLLER–LYER SCORES

	Mean	σ
ZAT	370.13	91.20
ZAP	98.50	45.67
ZBD	364.87	101.20
ZBN	364.87	101.20
OATD	1.11	3.75
OATN	3.21	2.19
OAPD	0.08	2.42
OAPN	1.84	1.56
OBD	1.15	4.30
OBN	3.54	2.66
LATD	−6.16	5.34
LATN	6.58	4.80
LAPD	0.35	1.92
LAPN	1.50	1.24
LBD	−6.53	5.34
LBN	6.91	4.83
SATD	0.35	3.99
SATN	3.18	2.40
SAPD	−0.04	2.59
SAPN	2.06	1.54
SBD	0.21	4.43
SBN	3.39	2.83
GATD	−0.25	3.46
GATN	2.74	2.10
GAPD	0.11	2.52
GAPN	1.93	1.60
GBD	−0.26	4.18
GBN	3.11	2.77
HATD	−0.41	3.21
HATN	2.49	2.05
HAPD	0.09	2.06
HAPN	1.68	1.18
HBD	−0.38	3.35
HBN	2.55	2.18
IATD	−0.55	3.47
IATN	2.49	2.46
IAPD	0.10	2.36
IAPN	1.17	1.60
IBD	−0.60	3.44
IBN	2.55	2.36
JATD	−1.00	3.88
JATN	2.85	2.79
JAPD	−0.07	2.25
JAPN	1.66	1.51
JBD	−1.09	4.15
JBN	3.03	3.01

OATD vs *OBD*, .92; *LATD* vs *LBD*, .95; etc.). Analogously, the correlations between the nondirectional bias scores and the corresponding nondirectional measures of accuracy from the true standard were also quite high (e.g., the correlation between *ZAT* and *ZBN* was .99; *OATN* vs *OBN*, .86; *LATN* vs *LBN*, .94; etc.).

The intercorrelations among the seven additional variables [diagnosis (psychotic vs normal), sex, age in months, plus time in seconds, and error measures from the BGT and the TMT] are also available from the author. Time was highly negatively related to error rate on the TMT ($r = -.93$), much less so on the BGT ($r = -.26$). Interestingly, diagnosis was highly positively associated with BGT error rate ($r = .71$), while it was negatively related to TMT error rate ($r = -.43$); while times on both tests were positively related ($r = .52$), the two error rates were negatively related to each other ($r = -.52$). Sex and age were not significantly associated with any of the BGT or TMT scores.

Table 6 lists the significant correlations between the MLI scores and those from the seven additional variables. Well over one-quarter of the total set of correlations between these measures were significant at the .05 level. TMT time, TMT errors, BGT errors, and diagnosis were significantly related to a host of MLI scores. Among the highest of these relations were those between TMT time and HATN ($r = .65$) or HBN ($r = .64$), and between diagnosis and ZAP ($r = .51$) or OAPN ($r = .47$). The high correlation between diagnosis and ZAP deserves some scrutiny. For while the *mean* responses of the average child in the psychotic and the normal samples were virtually identical (indeed, the correlation between diagnosis and the total accuracy measure, ZAT, was .00), there was a substantial difference between the samples in intraindividual variability about each subject's mean. *ZAP,* which reflects the amount of such intraindividual variability, is highly related ($r = .92$) to the *range* of responses for each subject (that is, the subject's highest setting among all the trials minus his/her lowest setting). In these samples, there were *no* normal children with range values higher than the median value for the psychotic.

An investigator who computed only the global accuracy measure, *ZAT*, would have been led to conclude that *MLI* performance was but weakly related to measures of speed and accuracy on the TMT, and completely unrelated to either BGT performance or to clinical diagnosis. Once again, however, the employment of a more articulated scoring procedure served to uncover some highly significant and clinically-important relationships, relationships that were not manifested in the composite measure of global accuracy. On the other hand, as with the RFT analyses presented earlier, the use of 46 MLI scores certainly served to increase the probability that some of the significant relationships displayed in Table 6 are unique to this sample, and consequently a replication of this study in another sample is now clearly needed.

TABLE 6
SIGNIFICANT CORRELATIONS BETWEEN THE MÜLLER–LYER SCORES AND THE SEVEN ADDITIONAL VARIABLES ($N = 60$)

				Bender–Gestalt		Trail Making	
	Diagnosis	Sex	Age	Time	Errors	Time	Errors
ZAT						.33**	−.38**
ZAP	.51**				.50**	.43**	−.42**
ZB						.29*	−.34**
OATD					.26*		
OAPN	.47**				.44**		
OBD					.26*		
LATN						.39**	−.37**
LAPD		.29*		−.26*			
LAPN	.31*	.25*					
LBN					.27*	.38**	−.35**
SATN	.30*	.25*					
SAPD						.27*	
SAPN	.38**				.43**	.36**	−.36**
SBN	.33**				.27*		
GATD		−.28*					
GATN	.34**				.41**	.32**	−.38**
GAPN			.25*				
GBD		−.29*					
GBN	.40**				.47**	.26*	−.29*
HATD						.39**	−.31*
HATN	.39**				.34**	.65**	−.62**
HBD						.37**	−.29*
HBN	.40**				.33**	.64**	−.60**
IATD						−.27*	
IATN	.33**				.33**	.34**	−.28*
IAPN					.31*		
IBD						−.29*	
IBN	.38**				.34**	.37**	−.31*
JAPN	.42**					.30*	−.34**
JBN	.28*						

* $p \leq .05$.
** $p \leq .01$.

GENERALIZING THE ANALYTIC SCHEMA

The framework which has now been illustrated with the Rod-and-Frame Test and the Müller–Lyer illusion should be generally applicable to most other objective tests, specifically whenever (a) the initial measures of performance consist of deviations from some objective or "true" standard, and (b) two or more experimental factors are represented across the test trials, ideally in a complete factorial design. For such cases, the present analytic schema can be viewed as an extension of standard analysis of variance methodology.

TABLE 7

AN ILLUSTRATIVE DESIGN FOR ANALYSES WITH THREE FACTORS AND TWO LEVELS PER FACTOR

Trial	Stimulus configuration			Total	Coding vectors						
No.	O	L	S	score (Z)	Factor O	Factor L	Factor S	$O \times L$ (G)	$O \times S$ (H)	$L \times S$ (I)	$O \times L \times S$ (J)
1	+	+	+	+1	+1	+1	+1	+1	+1	+1	+1
2	−	+	+	+1	−1	+1	+1	−1	−1	+1	−1
3	−	+	+	+1	−1	+1	+1	−1	−1	+1	−1
4	+	+	+	+1	+1	+1	+1	+1	+1	+1	+1
5	+	−	+	+1	+1	−1	+1	−1	+1	−1	−1
6	−	−	+	+1	−1	−1	+1	+1	−1	−1	+1
7	−	−	+	+1	−1	−1	+1	+1	−1	−1	+1
8	+	−	+	+1	+1	−1	+1	−1	+1	−1	−1
9	+	−	−	+1	+1	−1	−1	−1	−1	+1	+1
10	−	−	−	+1	−1	−1	−1	+1	+1	+1	−1
11	−	−	−	+1	−1	−1	−1	+1	+1	+1	−1
12	+	−	−	+1	+1	−1	−1	−1	−1	+1	+1
13	+	+	−	+1	+1	+1	−1	+1	−1	−1	−1
14	−	+	−	+1	−1	+1	−1	−1	+1	−1	+1
15	−	+	−	+1	−1	+1	−1	−1	+1	−1	+1
16	+	+	−	+1	+1	+1	−1	+1	−1	−1	−1

One begins with a three-way data matrix of size $3 \times k \times N$, where N is the number of subjects under study, k is the number of trials, and each of the three columns represent one of the three[1] types of initial scores: (C_i) the original *signed* deviations from the "true" standard, (M_i) the absolute values of C_i, and (Q_i) the absolute values of the deviations from the subject's mean across all trials ($\bar{C}$). For each subject in turn, the $3 \times k$ matrix of initial scores is postmultiplied by a matrix of coding vectors for each of the experimental effects and their interactions to produce derived scores of the sort displayed in Tables 1 and 4.

Two variants of this schema can be differentiated, the first where there are only two levels of each experimental factor (e.g., "earlier" vs "later" trials), and the second where there are three or more levels of at least one such factor (e.g., "early" vs "middle" vs "late" trials). When there are only two levels for all of the factors, "contrast coding" (Cohen & Cohen, 1975, p. 195) with +1 and −1 values in each coding vector will generate the same type of scores here analyzed for the RFT and MLI. When there are three or more levels for any factor, then the coding is slightly more complicated. In either case, each of the three $k \times N$ matrices of initial scores can be subjected to a repeated measures analysis of variance to ascertain the statistical significance of the main effects and their interactions, especially of all of the interactions involving the "subjects" factor.

Let us assume that there are two experimental factors, each measured at two levels, plus a "trial sequence" factor also of two levels ("earlier" vs "later" trials). An investigator who wished to have replications at the highest level of interaction would need to administer at least 16 trials, as illustrated in Table 7. This design is similar to that used in the Müller–Lyer study, so for convenience the factors in Table 7 are labeled as in Table 4. The columns labeled "stimulus configuration" indicate how each trial is constructed; positive and negative signs are used to design the two levels of each factor. The eight coding vectors provide the weights by which the initial score vectors are postmultiplied to obtain each of the eight types of derived scores. Since three types of initial scores are analyzed (C_i, M_i, and Q_i), there are 3×8, or 24, derived scores at this stage—namely all of the *directional* scores. Their absolute values provide the analogous *nondirectional* scores. This "complete set" of 48 derived scores may be culled to eliminate one from each of two pairs of redundant measures; specifically, the "directional" and "nondirectional" total accuracy scores are identical, since the sum of absolute values (M_i and Q_i) must always be positive. The remaining 46 derived scores are precisely those listed in Table 4.

[1] Note that the fourth logical type of data vector, the E_i values, are perfectly correlated with the C_i values—the former differing from the latter only by a constant (the subject's mean response, $\bar{C}$). In computing each of the final scores, that constant drops out (since it is simultaneously added to and subtracted from the contrasted trials), and thus the resulting scores based on the E_i values are identical to those based on the C_i ones.

REFERENCES

Asch, S. E., & Witkin, H. A. Studies in space orientation. I. Perception of the upright with displaced visual fields. *Journal of Experimental Psychology*, 1948, **38**, 325–337. (a)

Asch, S. E., & Witkin, H. A. Studies in space orientation. II. Perception of the upright with displaced visual fields and with body tilted. *Journal of Experimental Psychology*, 1948, **38**, 455–477. (b)

Benfari, R., & Vitale, P. Relationship between vertical orientation in the Rod and Frame Test and in a compensatory tracking task. *Perceptual and Motor Skills*, 1965, **20**, 1073–1080.

Blake, R. R., & Ramsey, G. G. V. (Eds.), *Perception: An approach to personality*. New York: Ronald Press, 1951.

Burtt, H. E. Educational research and statistics: Measuring interest objectivity. *School and Society*, 1923, **17**, 444–448.

Cabe, P. A. Note on response sets on the Rod-and-Frame Test, *Perceptual and Motor Skills*, 1968, **26**, 94. (a)

Cabe, P. A. The relation between the Rod-and-Frame Test and Witkin's Embedded Figures Test. *Educational and Psychological Measurement*, 1968, **28**, 1243–1245. (b)

Cattell, R. B., & Warburton, F. W. *Objective personality and motivation tests: A theoretical introduction and practical compendium*. Urbana, Illinois: Univ. of Illinois Press, 1967.

Coan, R. W. Factors in movement perception. *Journal of Consulting Psychology*, 1964, **28**, 394–402.

Cohen, J., & Cohen, P. *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, New Jersey: Lawrence Erlbaum, 1975.

Cronbach, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin*, 1955, **52**, 177–193.

Eysenck, H. J., & Slater, P. Effects of practice and rest on fluctuations in the Müller–Lyer illusion. *British Journal of Psychology*, 1958, **49**, 246–256.

Gage, N. L., & Cronbach, L. J. Conceptual and methodological problems in interpersonal perception. *Psychological Review*, 1955, **62**, 411–422.

Goldberg, L. R. On the analytic decomposition of Rod-and-Frame Test performance: An empirical comparison between global and component scores. *Oregon Research Institute Research Bulletin*, 1972, **12** (14).

Gruen, A. A critique and re-evaluation of Witkin's perception and perception-personality work. *Journal of General Psychology*, 1957, **56**, 73–93.

Hettema, J. Cognitive abilities as process variables. *Journal of Personality and Social Psychology*, 1968, **10**, 461–471.

Hundleby, J. D., Pawlik, K., & Cattell, R. B. *Personality factors in objective test devices: A critical integration of a quarter of a century's research*. San Diego: Knapp, 1965.

Klein, G. S. *Perception, motives, and personality*. New York: Knopf, 1970.

Krathwohl, D. R., & Cronbach, L. J. Suggestions regarding a possible measure of personality: The Squares Test. *Educational and Psychological Measurement*, 1956, **16**, 305–316.

Mischel, W. *Personality and assessment*. New York: Wiley, 1968.

Ormanli, M. The comparative effectiveness of the Bender–Gestalt and the Trail Making tests in differentiating psychotic from normal children. *Noro Psikiyatri Arsivi (Neuropsychiatry Archives)*, 1975, **12**, 9–23.

Ormanli, M. The Müller–Lyer illusion in psychotic and normal children. *Noro Psikiyatri Arsivi (Neuropsychiatry Archives)*, in press.

Pressey, A. W. Field dependence and susceptibility to the Poggendorff illusion. *Perceptual and Motor Skills*, 1967, **24** 309–310.

Robinson, J. O. *The psychology of visual illusion*. London; Hutchinson, 1972.

Thurstone, L. L. *A factorial study of perception*. Chicago: Univ. of Chicago Press, 1944.

Travis, L. E. A test for distinguishing between schizophrenoses and psychoneuroses. *Journal of Abnormal and Social Psychology,* 1925, **19,** 283–298.

Vernon, P. E. The distinctiveness of field independence. *Journal of Personality,* 1972, **40,** 366–391.

Wallace, J. An abilities conception of personality: Some implications for personality measurement. *American Psychologist,* 1966, **21,** 132–138.

Witkin, H. A. The nature and importance of individual differences in perception. *Journal of Personality,* 1949, **18,** 145–170. (a)

Witkin, H. A. Perception of body position and of the position of the visual field. *Psychological Monographs,* 1949, **63,** 7 (Whole No. 302). (b)

Witkin, H. A., & Asch, S. E. Studies in space orientation: III. Perception of the upright in the absence of a visual field. *Journal of Experimental Psychology,* 1948, **38,** 603–614. (a)

Witkin, H. A., & Asch, S. E. Studies in space orientation: IV. Further experiments on perception of the upright with displaced visual fields. *Journal of Experimental Psychology,* 1948, **38,** 762–782. (b)

Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. R., & Karp, S. A. *Psychological differentiation: Studies of development.* New York: Wiley, 1962.

Witkin, H. A., Lewis, H. B., Hertzman, M., Machover, K., Meissner, P. B., & Wapner, S. *Personality through perception: An experimental and clinical study.* New York: Harper, 1954.

Zigler, E. A measure in search of a theory? (Review of H. A. Witkin, et al., *Psychological differentiation: Studies of development*). *Contemporary Psychology,* 1963, **8,** 133–135.