

## THE PREDICTIVE VALIDITY OF GRAPHOLOGICAL INFERENCES: A META-ANALYTIC APPROACH\*

EFRAT NETER and GERSHON BEN-SHAKHAR

The Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel

(Received 12 July 1988)

**Summary**—The use of graphology as a device for personnel selection is prevalent and increasing. This study attempts to examine the validity of graphology in this particular applied field by means of meta-analysis—a method of integrating research findings across studies. Seventeen studies dealing with the validity of graphology as a personnel selection device were tracked down. A total of 63 graphologists and 51 nongraphologists who evaluated 1223 scripts were included in the data set. The group of non-graphologists served as a control group to establish a base line for the predictive validity that could be obtained on the basis of the script's content without the benefit of any graphological knowledge. Representative correlations between handwritten—based inferences and criteria were calculated for each judge in each study, and correction for sampling error was performed. It was found that correlations between inferences based on content-laden scripts and a general criterion range from 0.136 to 0.206 for all judges (graphologists, psychologists, laymen) and 0.153–0.177 for graphologists. In the few cases where neutral scripts were used the validities of the graphologists were near zero. In addition it was found that psychologists (with no knowledge in graphology) outperformed graphologists on all dimensions. The results were discussed, suggesting that the source of the limited validity of handwriting analysis may be the script's content.

### INTRODUCTION

Graphology, a method of inferring personal attributes through handwriting analysis, is most widely applied in the sphere of personnel selection. It is used as a screening tool that helps determine the compatibility of a person to a particular job. Levy (1979) reports that 85% of European firms use graphology; Klimosky and Refaeli (1983) report that 3000 American firms employ it, and according to Ben-Shakhar, Bar-Hillel, Bilu, Ben-Abba and Flug (1986) graphology is more widespread in Israel than any single personality test.

Most research on graphology relates to three main issues: the reliability of handwriting, the reliability of the interpretations, and the validity of the inferences. The reliability of handwriting relates to the consistency and the stability (over time and changing external circumstances) of various features of the handwriting of a given writer. The reliability of the interpretations includes inter-judge reliability (i.e. degree of agreement between different graphologists interpreting the same scripts), test–retest reliability (i.e. consistency of judgements made by the same graphologist across different samples of handwriting of the same writer), and the effect of content-laden scripts on the interpretation. The validity of graphological inferences refers to their success in predicting future behavior.

Four reviews (Fluckinger, Tripp and Weinberg, 1961; Lockowandt, 1976; Klimoski and Refaeli, 1983; Nevo, 1986a) have presented studies dealing with the reliabilities and validities of judgments based on handwriting analysis. However, these reviews do not provide comprehensive conclusions regarding the different aspects of reliability and validity. While the evidence supports a relatively high degree of stability in people's handwriting (i.e. the test–retest correlation coefficients of various features of handwriting are usually 0.70 or higher), reports about the reliabilities of the interpretations are much less uniform. For example, in a study by Hofsommer, Holdsworth and Seifert (1965) leadership ability among foremen was rated by three graphologists with an inter-judge agreement of  $r = 0.74$ ; on the other hand, Lockowandt (1976) cites a study that obtained an average agreement of only  $r = 0.39$  between three graphologists assessing the success of technical school students on the basis of their handwriting. Lockowandt (1976) presents studies

\*This article is based on a master's thesis conducted by the first author under the direction of the second author. We thank Baruch Nevo and Oscar Lockowandt for their help in tracing the studies upon which the meta-analytic work is based.

that looked into the judgement of a single graphologist across different samples of the same writer and obtained correlations of 0.78–0.88 (e.g. Crider, 1941; Reichold, 1969). Nevo (1986b), in his review on the reliability of graphology, divides the reliability into three categories; graphometric measures (e.g. letter size, letter shape, margin width, space between lines), graphoimpressionistic characteristics (e.g. roundness, rhythms, pressure, decorativeness) and graphodiagnostic scales (e.g. ego strength, emotional stability, etc). He concludes that the range for most of the reported reliabilities of graphometric measures is approx. 0.70–0.90; the reliability range for the majority of graphoimpressionistic characteristics is 0.40–0.80, and the corresponding range for graphodiagnostic scales is the lowest, 0.30–0.60. The effect of content, as Klimosky and Refaeli note (1983), has been generally overlooked. Few studies with neutral scripts were conducted, and even fewer examined the issue systematically. Klimosky and Refaeli cite their own study (1983), in which they found that content-laden scripts had little effect on the graphologists assessments. Ben-Shakhar *et al.* (1986), on the other hand, have suggested that script content and aesthetics could account for the small validity coefficients obtained by graphologists and other people when making predictions on the basis of hand-written scripts.

The evidence on the critical issue of validity is even less definitive. For example, while Hofsommer, Holdsworth and Seifert (1962) obtained correlations as high as 0.55 between predictions made by graphologists and the criterion of supervisors assessments, other studies (e.g. Strolowitz, 1980; Wallner, 1963) found near zero validities for graphological judgments against similar criteria. The reviews cite studies using different criteria, employing varied procedures and yielding very different results. The conclusions of the reviewers are quite similar. Fluckinger *et al.* (1961) conclude that the research “on the whole [is] fragmentary” (p. 86); Klimoski and Refaeli (1983) note that “definitive research is not yet available” (p. 200); Lockowandt (1976) writes that “the validity of the variables has yielded up to now no convincing success” (abstract). Concurrently, the reviewers call for more “sophisticated research” and for “further scientific clarification”.

While these recommendations are certainly welcome, it is doubtful whether additional studies would yield uniform conclusions. The lack of uniformity may be the consequence of the heterogeneity of the methods employed by the different graphologists and the different criteria and circumstances under which the handwriting analyses were conducted, but it could result, at least to some extent, from random variations between studies and from sampling error. An approach attempting to provide a quantitative summary of several studies, and to discriminate real from artificial variability between studies was developed by Hunter, Schmidt and Jackson (1982). Their technique, as well as other meta-analytic techniques that were promoted in the last decade (e.g. Glass, 1976) call for making more sense of the available data rather than collecting more data. The present study applies the meta-analytic techniques developed by Hunter *et al.* (1982) to estimate the validity of predictions based on handwriting analysis. This technique is a quantitative cumulation of results across studies that is accompanied by the correction of sampling errors, errors of measurement, and range restriction. The outcome is usually a stable pattern, unattainable when single studies are examined and the reviewers rely on statistical significance tests to evaluate them. Meta-analysis is particularly suited for examining the validity of graphology because single studies cannot represent the great variety of graphological methods. Furthermore, most of the studies use a small number of Ss (i.e. both writers and graphologists) and sometimes even a single graphologist. This makes generalizing and drawing conclusions from single studies difficult, and hence there is a strong need for accumulating data across studies.

Nevo (1986a) has recently accumulated validity studies on graphology conducted in Israel. He computed a general estimate for the validity of graphological predictions by weighting a representing correlation of each study by its sample size. The weighted average result of this computation was a correlation of  $R = 0.14$ . However, Nevo's study is limited from several perspectives: (1) it is confined only to studies conducted in Israel; (2) it does not include an attempt to estimate the variance due to the sampling error, and to discriminate between real and artificial variance between studies; (3) it does not include the essential control of non-graphologists trying to make the same predictions as graphologists. We find this comparison between the validities produced by graphologists and by lay persons essential, because content-laden scripts (e.g. life histories) were used in most studies. Such scripts contain a great deal of non-graphological information (e.g. education, previous work records) regarding the writer. Previous studies

(Ben-Shakhar *et al.*, 1986; Jansen, 1973) have demonstrated that this kind of information could be utilized to produce positive validities. The use of a control group of non-graphologists may help, to some extent, in controlling against the alternative hypothesis that predictions are based on the biographical information rather than on the graphological features of the scripts.

The purpose of the present study is to collect all the available studies on the prediction of future job performance by graphology, made by both professional graphologists and lay persons. Correlations between these judgements and external criteria will provide the data-base upon which the meta-analysis will be carried out. The validity coefficients of graphologists and lay persons will be compared. The residual variance remaining after subtracting the sampling error variance from the observed variance will provide an estimate for the effect of real differences among graphologists. Possible moderating variables that may contribute to the variance in the results will be considered. Such moderators may include the nature of the criterion, the type of the script (content-laden or not), and the type of job.

## METHOD

### *Compilation of data*

A search for published and unpublished studies relating to graphological inferences in the setting of personnel selection was undertaken. To find studies of graphology we carried out manual searches of two library data bases: Comprehensive Dissertation Abstracts and Psychological Abstracts. The bibliographies and articles located through these data bases provided second sources of studies for meta-analysis. We also approached researchers whom we knew to be active in the field and asked for bibliographies and studies.

We collected data only from studies that met the following requirements: (a) they reported validity results in the form of correlation coefficients or raw data that contained sufficient information to compute a correlation coefficient, (b) the validity estimates were based on an external criterion such as job proficiency appraisal or training success, (c) the reporting of sample size. Examples of studies that were not included are a study that looked into the predictive validity of graphological classification of integrity but reported only significance levels of the statistical tests (M.M. and A. G., 1985), and a study that examined the validity of graphological and psychological inferences for predicting adjustment to a small community, but lacked a sound measure of this criterion (Shilo, 1979). The collected data sets included 17 studies: 11 published papers and 6 unpublished. The unpublished papers were M.A. theses or internal research reports; all the unpublished data originated in Israel.

The studies were classified according to the following items: type of criterion (supervisory appraisal, production data, grades in training), type of correlation, number of scripts and their type (content-laden or neutral), type of judge (graphologist, psychologist, layman) and their number, and the representing correlation coefficient between the inferences and the criteria. In some of the studies the representative correlation is the value reported in the article; in others, it was necessary to calculate the representative correlation from the reported data or to transform the original statistic into another type of coefficient to allow comparisons across studies. The classification of the 17 studies is displayed in Table 1. A separate card file was established to record narrative background information on each study. The information included the original source of the study, its data, its authors, the population on which it was carried out, and its procedures.

As each study typically included many predicting variables with different labels, we categorized all variables into the following three dimensions: work proficiency, social-psychological attributes, and a general evaluation. All the reported coefficients were categorized into one of these dimensions by two judges, and whenever a disagreement occurred, a third judge made the decision. For some purposes it is desirable to represent each study by a single coefficient, hence we defined a 'General dimension' which was identical to the general evaluation dimension whenever the latter was initially included in the study. In other cases, the work-related dimension was used, and in the few cases where both the general evaluation and the work-related dimensions were missing, the social-psychological dimension was used as the general dimension. Following the categorization, a median coefficient was computed for each judge across the different variables within each dimension.

Table 1. List of studies and their characteristics

Authors and year	Population	Criteria	Type of correlation	No. of scripts and its type		Type of judges and their number	Representing† correlation
				Neutral	Content		
(1) Sonnemnen and Kerman (1962)	Executives	Supervisory ratings	Weighted Capa*	37		Graphologists—1	0.31
(2) Hofsommer, Holdsworth and Seifert (1962)	Training foresters	Grades by supervisors	Spearman	54		Graphologists—2	0.55
(3) Hofsommer and Holdsworth (1963)	Training pilots	Training success determined by supervisors	Point biserial*	141		Graphologists—1	0.20
(4) Wallner (1963)	Training executives	Supervisory ratings	Pearson	89		Graphologists—2	0.05
(5) Zdep and Weaver (1967)	Insurance sales Personnel	Initial year commission on life insurance	Pearson	63		Graphologists—2	0.08
(6) Jansen (1973) (Study 3)	Commercial—contact personnel	Supervisory ratings	$\phi^*$	20		Graphologists—10	0.45
(7) Jansen (1973) (Study 4)	Commercial—administration personnel	Supervisory ratings	Kendel $\tau^*$	9		Psychologists—10	0.45
(8) Strolovitz (1980)	Scientific—technical personnel	Supervisory ratings	Pearson	25		Laypersons—10	0.50
(9) Flug (1981)	Bank personnel	Supervisory ratings	Pearson	58		Graphologists—6	-0.02
(10) Rafael and Klimosky (1983)	Real estate salesmen	Supervisory ratings	Pearson	70		Laypersons—6	-0.02
(11) Keinan, Barak and Ramati (1984)	Infantry cadets	Supervisory ratings; sales rating	Kendel coefficient of concordance	65		Graphologists—2	-0.19
(12) Esroni, Rolnik and Livnat (1985)	Military personnel	Supervisory ratings; actual success	$\phi^*$	23		Psychologists—3	0.21
(13) Esroni Rolnik and Livnat (1985)	Military personnel	Supervisory rating	$\phi^*$	125		Psychologists—1	0.28
(14) Esroni, Rolnik and Livnat (1985)	Military personnel	Supervisory rating; actual success	Pearson	49		Graphologists—20	0.18 (content)
(15) Esroni, Rolnik and Livnat (1985)	Military personnel	Supervisory rating	$\lambda^*$ —Predictive association for categorical data	45		Graphologists—6	-0.01 (neutral)
(16) Bornstein (1985)	Cadets	Actual success and course grades determined by supervisors	Pearson	214		Psychologists—6	0.26
(17) Drori (1986)	Technical Clerical	Supervisory ratings	Pearson	66		Laypersons—6	0.20
						Graphologists—1	0.11
						Graphologists—1	0.19
						Graphologists—1	0.01
						Graphologists—1	0.03
						Graphologists—1	0.06
						Graphologists—3	0.09
						Psychologists—3	0.12
						Laypersons—3	0.16
						Graphologists—1	0.42

†An asterisk appears whenever we computed the correlation.

‡The representing correlation is the median correlation across judges for the General dimension.

Table 2. Results of meta-analysis based on data-set A

Statistics dimension	$r$	Observed variance ( $s^2$ )	$n$ of coefficients	$Se^2$	Remaining variance
<b>Graphologists</b>					
Work	0.213	0.044	28	0.023	0.021
Social-psychological	0.143	0.013	18	0.019	0
General evaluation	0.155	0.022	17	0.011	0.011
General	0.177	0.033	40	0.017	0.016
<b>Psychologists</b>					
Work	0.292	0.070	18	0.040	0.030
Social-psychological	0.226	0.047	8	0.041	0.006
General evaluation	0.153	0.02	5	0.006	0
General	0.193	0.31	21	0.020	0.011
<b>Laypersons</b>					
Work	0.296	0.087	17	0.049	0.038
Social-psychological	0.058	0.016	7	0.058	0
General evaluation	0.152	0.000	4	0.005	0
General	0.206	0.033	20	0.020	0.013
<b>Neutral scripts (by graphologists)</b>					
Work	-0.010	—	1	0.014	—
Social-psychological	0.038	0.002	2	0.015	0
General	0.033	0.002	2	0.015	0

Since some studies included several judges within each type (graphologists, psychologists and laymen) and reported results for each judge, we compiled two data-sets: one in which the data of each judge in each study was separately reported (hereafter: data-set A), and another in which each type of judge within each study was represented by a single coefficient for each dimension—the median computed across judges within each type and each dimension (hereafter: data-set B).

## RESULTS

A meta-analysis was conducted for each type of judge (graphologist, psychologist, layperson) and for each dimension (work, social-psychological, general evaluation, and General). It was carried out on the two aforementioned data-sets: (1) data-set A, where the correlations of the individual judges were averaged across studies (whenever there was more than one judge and according to the type of judge). The mean coefficients (weighted by the number of scripts) for each dimension are presented in Table 2, along with the corresponding observed variances, the number of coefficients on which the computations were based, the sampling error, and the variance remaining after subtracting the sampling error variance from the observed variance; (2) data-set B, where each dimension in each study is represented by a single correlation for each type of judge. The meta-analytic statistics for data-set B are presented in Table 3. All the statistics were computed using the procedures and formulae described in Hunter *et al.* (1982).<sup>\*</sup> No corrections for restriction of range and for error of measurement were performed, as the necessary data were unavailable in most of the studies. Still, their exclusion was not deemed critical; Schmidt, Hunter and Caplan (1981, p. 266) report that most of the variance due to artifacts is accounted for by the sampling error, which constitutes 75% of the artificial variance in each study, and 90% across studies.

Table 2 shows that the coefficients for the General dimension range between 0.177 and 0.206 for all judges. The only exception is the case of neutral scripts analyzed by graphologists, where the mean correlation is 0.033. The range of these correlations in Table 3 is 0.136–0.180.

The most surprising finding arises when the different judges are compared: it appears that graphologists are less successful in predicting future behavior from handwriting analysis than psychologists in almost all the dimensions (in both data-sets); graphologists and laypersons alternate in their success pattern across the different dimensions. Even more intriguing is the finding

<sup>\*</sup>The correlations in data-set B are based on more than one judge, therefore their sampling error variance may be smaller than the values computed on the basis of the sample size (the number of the scripts alone). However, as a formula for the variance of the median correlation is not available, we estimated each error variance twice: (1) the sample size used was the number of scripts, as in the case of data-set A; (2) the sample size used to compute the error variance for each correlation was the number of scripts multiplied by the number of judges. Only the values computed by the first method are presented in Table 3. The second method yielded smaller values for the estimated sampling errors, and hence larger values for the remaining variances, but it did not have an effect on the pattern of the results or on the conclusions drawn from it.

Table 3. Results of meta-analysis based on data-set B

Statistics dimension	$\bar{r}$	Observed variance ( $s^2$ )	$N$ of coefficients	$Se^2$	Remaining variance
<b>Graphologists</b>					
Work	0.206	0.022	8	0.017	0.005
Social-psychological	0.134	0.012	7	0.016	0
General evaluation	0.144	0.025	10	0.012	0.013
General	0.153	0.026	16	0.014	0.012
<b>Psychologists</b>					
Work	0.256	0.016	4	0.023	0
Social-psychological	0.252	0.023	3	0.020	0.003
General evaluation	0.169	0.003	3	0.008	0
General	0.180	0.008	5	0.013	0
<b>Laypersons</b>					
Work	0.107	0.022	4	0.024	0
Social-psychological	-0.004	0.001	3	0.021	0
General evaluation	0.019	0.004	3	0.009	0
General	0.136	0	5	0.013	0
<b>Neutral scripts (by graphologists)</b>					
Work	-0.010	0.000	1	0.014	—
Social-psychological	0.038	0.002	2	0.015	0
General	0.033	0.002	2	0.015	0

that judges (graduate students of psychology), using printed text, did better than graphologists. However, this finding is based on only one study (Jansen, 1973), and its procedure was never replicated.

The results of Tables 2 and 3 indicate that although the variability was reduced by correcting for sampling error, it was not eliminated in most cases (see the right hand columns of Tables 2 and 3). There are two possible reasons for the remaining variability among judges: (1) The remaining variability is artifactual, and it is due to factors not corrected for in the present analysis (e.g. range restriction, errors of measurement). (2) There is some true variability among judges, and although generally graphologists are poor predictors of performance, some of them are better than others, indicating that in principle there might be some validity in handwriting analysis. It must, however, be kept in mind that this point is true for non-graphologists as well. In other words, the remaining variability could mean that some people are better than others in their ability to make predictions on the basis of content-laden scripts, and this assertion has little to do with what graphology is claiming to be doing.

We have tried to further decrease the variance by looking for moderating variables. One possibility is the nature of the scripts (i.e. content-laden vs neutral). Indeed, a considerable difference exists: the average correlations for content-laden scripts are 0.177 and 0.153 in data-sets A and B, respectively, as opposed to 0.033 for neutral scripts (see Tables 2 and 3). However, the latter result is based on only two studies, and is therefore of limited generalizability.

Another possible moderating variable is the setting in which the study was conducted. Graphologists, and judges in general, may differ in their predictive ability in different settings. The relevant differentiation in our data is work setting vs training setting. We therefore divided data-set A into work vs training setting. The results for these two settings, with relation to the general dimension, are presented in Table 4.

These results indicate that in the training setting the variability was considerably reduced for all types of judges, and in the case of the non-graphologists eliminated. On the other hand, the

Table 4. Results of meta-analysis conducted separately for work and training setting (data-set A)

Statistics Judge type	$\bar{r}$	Observed variance ( $s^2$ )	$n$ of coefficients	$Se^2$	Remaining variance
<b>Work setting</b>					
Graphologists	0.246	0.048	25	0.031	0.017
Psychologists	0.308	0.080	17	0.045	0.035
Laypersons	0.354	0.093	16	0.048	0.045
<b>Neutral scripts (by graphologists)</b>					
Training setting	0.033	0.002	2	0.015	0
Graphologists	0.145	0.023	15	0.009	0.014
Psychologists	0.142	0.001	4	0.005	0
Laypersons	0.152	0.000	4	0.005	0
Neutral scripts	—	—	—	—	—

variability was not reduced in the work setting where the factors determining success are more numerous and complex. The average correlations between the inferences and the criteria are higher in the work setting than in the training setting, and the success pattern among the judges is different: in the work setting the laypersons did best, followed by the psychologists and the graphologists, while in the training setting the laypersons are followed by graphologists and then psychologists. At any rate, the division into different settings did not eliminate the variance, nor did division into studies carried out in Israel as opposed to studies conducted in other countries yield any significant decrease in the variance. Again, note that corrections for range restriction and error of measurement were not performed, and some of the remaining variance may be accounted for by these artifacts.

## DISCUSSION

The results of this meta-analysis show that graphologists are not better than non-graphologists in predicting future performance on the basis of handwritten scripts. Actually, the graphologists' predictions had somewhat lower correlations with the criteria than those of non-graphologists. The graphologists' results were much better when they analyzed content-laden material than when they used neutral scripts, but this finding is based on too few studies to allow generalizations. The results also indicate that not all the variance between the judges was accounted for by sampling error. The variance was found to be higher in the work-setting than in the training-setting, indicating that the factors that determine success in the former setting may be more complex in their operation.

The implications of the first finding—i.e. graphologists cannot better predict future performance by means of handwriting analysis any better than non-graphologists—is intriguing. The comparison of non-graphologists to graphologists constitutes a control procedure that may provide a baseline for evaluating the performance of graphologists. From this point of view, the results discourage the use of graphology as a predictive tool.

On the other hand, the possibility that graphologists and non-graphologists derive their inferences from different sources cannot be excluded. There appear to be two avenues to examine the issue: the first is to assess the contribution of each graphological cue to the validity of the inference, and the second is to examine more closely the impact of content-laden vs content-free scripts on the validity of the inferences. Some work has already been done in both directions. As to the first, research on the atomistic aspects of graphology has proved to be unfruitful; Fluckinger *et al.* (1961) note that discrete and individual signs, or cues, are seldom thought of as having stable meaning, but rather vary depending upon the context in which they appear. Moreover, we could not find any serious theoretical account relating graphological signs to either personality traits or behavior. Graphological analysis is an attempt to infer from how people behave in one context to what kind of people they really are. It relies on the very general assumption that the characteristics of behaviour in that single context, as expressed by handwriting features, are indicative of the personality as a whole, and therefore of the entire range of an individual's behaviour. Beyond this general assertion the linkage between traits such as honesty, leadership, or intelligence and specific features of handwriting is not specified or explained, and there is no reason to believe that such a correspondence exists. This lack of theoretical background for the inferences made from handwriting to personality and behavior increases the likelihood that the small correlations that were obtained for the handwriting—based judgements reflect the value of the information contained in the scripts, rather than the validity of pure graphological signs.

As to the second direction, it has not been studied extensively: very few studies employed a design that examined the performance of graphologists and non-graphologists using content-free scripts. The results thus far suggest that graphologists may indeed make use of the content of the script; their performance decreases significantly when they do not use content-laden scripts. Ben-Shakhar *et al.* (1986) have recently reported an additional study that utilized neutral scripts. The study was not included in the present meta-analysis as it was not conducted in the personnel selection context, and it used a different criterion (five graphologists were asked to guess, on the basis of neutral handwritten scripts, which of eight professions is the writer's true profession). The task was performed by graphologists at a level that was not significantly different than chance, thus increasing the likelihood of the content interpretation of the graphologists' performance.

A definite conclusion as to the role of content on the judgments made by graphologists maybe premature at the present stage, as the available data is too scarce. Yet, this possibility cannot be ruled out, and further research attempting to establish the validity of pure graphological information should either demonstrate that graphologists can outperform laypersons in their predictions, or provide stronger evidence for the validity of graphological inferences without the benefit of context or any other non-graphological information.

The remaining variance among the judges, which might be construed as indicating the existence of real differences in ability among graphologists, can be accounted for by several explanations. First, though we have eliminated the largest source of error—sampling error—it is only one of the eight potential sources of variance in the distribution of validity coefficients cited by Pearlman, Schmidt and Hunter (1980). The list also includes variance due to differences in criterion reliability, test reliability, range restriction, criterion contamination and deficiency, computational, typographical and data recording errors, factor structure of tests measuring the same construct, and factor structure differences between criterion measures—none of which was dealt with. Second, the remaining variance is present not only among graphologists but also among the other types of judges—psychologists and laypersons. Hence, we are inclined to conclude that the variance is due mainly to remaining sources of error and not to meaningful differences in the predictive ability of the graphologists.

The results of the present study merge with directions pointed in the aforementioned reviews. Particularly worth mentioning is Nevo (1986a); his review of studies conducted in Israel, in many different settings, yielded an average validity of 0.14 for predictions made by graphologists. Such a validity, as noted by Nevo, has doubtful applicable implication. Our results indicate somewhat higher validities for the graphologists working with content-laden scripts (0.153–0.177) but show that non-graphologists, working with the same scripts, can achieve similar validities. Still, in spite of the discouraging findings, the use of graphology seems to be increasing, and the obvious question is—why? Bar-Hillel and Ben-Shakhar (1986) suggested several reasons for this unfounded popularity: (1) face validity which refers to the *appearance* of graphology as having the appropriate properties for reflecting personality; (2) personal validity which refers to the subjective feeling, imparted by exposure to graphology, that it is accurate and manages to capture the core of one's personality. Apparently, these factors make graphology attractive. Employers should be informed of the gap between the public impression and the limited predictive efficiency, now based on findings of accumulated data. Awareness of this gap may contribute to a more restricted use of handwritten scripts in determining a person's compatibility to a given profession or job.

## REFERENCES

- Bar-Hillel M. and Ben-Shakhar G. (1986) The *a priori* case against graphology. In *Scientific Aspects of Graphology* (Edited by Nevo B.). Thomas, Springfield, Ill.
- Ben-Shakhar G., Bar-Hillel M., Bilu Y., Ben-Abba E. and Flug A. (1986) Can graphology predict occupational success? Two empirical studies and some methodological ruminations. *J. appl. Psychol.* **71**, 645–653.
- Crider B. (1941) The reliability and the validity of two graphologists. *J. appl. Psychol.* **25**, 323–325.
- Fluckinger F. A., Tripp C. A. and Weiberg G. H. (1961) A review of experimental research in graphology, 1933–1960. *Percep. Mot. Skills* **12**, 67–90.
- Glass G. V. (1976) Primary, secondary, and meta-analysis of research. *Educ. Res.* **5**, 3–8.
- Hofsommer W., Holdsworth R. and Seifert T. (1965) Zur Reliabilitätsfragen in der Graphologie. *Psychologie Prax.* **9**, 14–24.
- Hunter J. E., Schmidt F. L. and Jackson G. B. (1982) *Meta-Analysis; Cumulating Research Findings Across Studies*. Sage, Beverly Hills, Calif.
- Jansen A. (1973) *Validation of Graphological Judgments; An Experimental Study*. Mouton, Paris.
- Klimosky R. J. and Refaeli A. (1983) Inferring personal qualities through handwriting analysis. *J. occup. Psychol.* **56**, 191–202.
- Levy R. (1979) Handwriting and hiring. *Dun's Review* March, 72–79.
- Lockowandt O. (1976) Present status of investigation of handwriting psychology as a diagnostic method. *J. Suppl. Abstr. Serv., Am. Psychol. Assess.*
- M. M. and A. G. (1985) Predictive validity of graphological classification of integrity. Internal unpublished research report by an Israeli financial institution.
- Nevo B. (1986a) Graphology validation studies in Israel: summary of 15 years of activity. A paper presented at the 21st International Congress of Applied Psychology, Jerusalem.
- Nevo B. (1986b) Reliability of graphology; a survey of the literature. In *Scientific Aspects of Graphology* (Edited by Nevo B.). Thomas, Springfield, Ill.
- Pearlman K., Schmidt F. L. and Hunter J. E. (1980) Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *J. appl. Psychol.* **65**, 373–406.



- Reichold L. (1969) Die Reliabilität und Validität graphologischer Aussagen. *Z. Menschenkunde* 33, 198–210.
- Schmidt F. L., Hunter J. E. and Caplan J. R. (1981) Validity generalization results for two groups in the petroleum industry. *J. appl. Psychol.* 66, 261–273.
- Shilo S. (1979) Prediction of success on a Moshav according to graphological scores as compared to prediction of the same criterion by psychological scores. Internal research report, Hadassa Institute for Career Guidance Counselling, Jerusalem, Israel.
- Strolovitz I. (1980) Impact of Personal Variables and Job Variables on the predictive Validity of Some Personnel Selection Practices for Scientific-technical Positions. M.A. thesis, the Technion.
- Wallner T. (1963) Über die Validität Graphologischer Aussagen. *Diagnostica* 9, 26–35.

## APPENDIX

### *Studies included in the Meta-Analysis of Graphological Inferences in the Context of Personnel Selection*

- Bornstein Y. (1985) Examination of the efficiency of graphology as a selection tool in the Israel Defence Forces. M.A. thesis, University of Haifa.
- Drori A. (1986) Graphology and job performance—a validation study. In *Scientific Aspects of Graphology* (Edited by Nevo B.). Thomas, Springfield, Ill.
- Esrone G., Rolnik A. and Livnat E. (1985) Studies Evaluating the Validity of Graphology in a Voluntary Military Unit. A paper presented at the 20th Israeli Psychological Association Conference.
- Flug A. (1981) A validation study of graphological evaluation in personnel selection. M.A. thesis, Hebrew University of Jerusalem.
- Hofsommer W. and Holdsworth R. (1963) Die Validität der Handschriftenanalyse bei der Auswahl von Piloten. *Psychol. prax.* 7, 175–178.
- Hofsommer W., Holdsworth R. and Seifert T. (1962) Zur Bewahrungskontrolle Graphologischer Diagnosen. *Psychol. Beit.* 7, 397–401.
- Jansen A. (1973) *Validation of Graphological Judgments; An Experimental Study*. Mouton, Paris.
- Keinan G., Barak A. and Ramati T. (1984) Reliability and validity of graphological assessment in the selection process of military officers. *Percept. Mot. Skills* 58, 811–821.
- Refaeli A. and Klimoski R. J. (1983) Predicting sales success through handwriting analysis: an evaluation of the effects of training and handwritten sample content. *J. appl. Psychol.* 68, 212–217.
- Sonnenman U. and Kerman J. P. (1962) Handwriting analysis—a valid selection tool? *Personnel* 39, 8–14.
- Strolovitch I. (1980) Impact of personal variables and job variables on the predictive validity of some personnel selection practices for scientific-technical positions. M.A. thesis, the Technion.
- Wallner T. (1963) Über die Validität Graphologischer Aussagen. *Diagnostica* 9, 26–35.
- Zdep S. M. and Weaver H. B. (1967) The graphoanalytic approach to selecting life insurance salesmen. *J. appl. Psychol.* 51, 295–299.