

Effects of Computerized Administration on Scores on the Minnesota Multiphasic Personality Inventory

Bruce H. Biskin
University of Maryland

Ronette L. Kolotkin
University of Minnesota

This study investigated the effects of administering a personality inventory by computer. Both the results of the initial study and a replication suggest that significant differences exist between paper-pencil and computer administrations of the MMPI on the cannot say (?) scale and scale 6 (Paranoia). However, there appears to be no set of items that would account for these scale differences. Differences on the ? scale were explained in terms of the different methods used to omit items in each condition. Differences on scale 6 were small, and the clinical significance of that difference needs to be investigated further. Implications for future research on computer-administered personality instruments are discussed.

During its relatively short existence, the digital computer has assumed an increasingly important role in the behavioral sciences. Computers have been used not only to score tests and inventories, but also to yield interpretations of personality inventories, such as the Minnesota Multiphasic Personality Inventory (MMPI) and the Sixteen Personality Factor Questionnaire (16 PF; e.g., Finney, 1969; Fowler, 1967; Rome, Swenson, Mataya, McCarthy, Pearson, & Keating, 1962). Research is currently underway to determine the efficacy and feasibility of administering ability tests on interactive computer sys-

tems (e.g., Weiss, 1975). A reasonable extension of computer-administered ability tests has been investigation into the administration of personality measures by computer.

Computer administration of psychological instruments may reduce response bias to certain kinds of items. Smith (1963) believed that, because of the impersonal nature of the computer, a testee would respond more honestly to "confessor-type" items (i.e., highly personal or revealing items) on a computer-administered test than on a paper-pencil test. He also noted that even if motivated to do so by exposure to later items, a test taker would be unable to modify previous responses on a computer-administered inventory. Thus, another source of response bias would be eliminated.

No evidence is available in the literature to support an alternative hypothesis that testees would tend to be less honest on a computer-administered test than on a paper-pencil test. However, this is not an unreasonable supposition, particularly if the computer is perceived as a huge data bank of highly personal information that may be accessed at any time without the knowledge or consent of the test taker.

If a paper-pencil inventory is to be modified for computer administration, there should be adequate evidence that the two modes of administration produce essentially the same scores. If differences are found, the investigator should demonstrate that the computerized test

yields information which is reliable and useful beyond that yielded by the paper-pencil form. This question has been addressed by Lushene, O'Neill, and Dunn (1972), who administered the MMPI twice to two groups of female undergraduates in a counterbalanced design. One MMPI was given in the standard paper-pencil mode, while the other was administered on a cathode-ray tube (CRT) computer terminal. Lushene et al. concluded that the two modes of administration yielded equivalent results on the basic scales and on a set of MMPI scales developed by Wiggins (1969). Lushene et al. neglected to report scores on the "cannot say" scale ("?"). This is an important omission, since leaving items blank can reduce scores on one or more scales, possibly flattening profiles and invalidating the results.

The present study was an initial attempt to determine the feasibility of developing a completely automated assessment program. Computer-administered personality tests are being used in research (Dunn, Lushene, & O'Neill, 1971; Lushene et al., 1972) and have already been incorporated into the admitting system of a mental health clinic (Johnson & Williams, 1975); however, the equivalence of the computerized forms of these inventories with the paper-pencil forms has not yet been adequately demonstrated for various clinical and non-clinical populations. The present studies attempted to clarify what differences, if any, exist between computerized and paper-pencil forms of the MMPI in a college population. Two experiments are reported: the first compares scores on computer-administered MMPIs with scores on paper-pencil MMPIs; the second is a partial replication of the first with several procedural changes.

Experiment 1

Method

Subjects. A total of 126 male undergraduates in the introductory psychology course at the University of Minnesota participated in this experiment.

Experimenters. Three graduate students in psychology served as the experimenters.

Materials and apparatus. The MMPI was administered in three modes: (1) in the standard paper-pencil form, using Hankes answer sheets; (2) by means of a Datapoint 3000 CRT terminal connected to the University of Minnesota CDC 6400 time-shared computer, transmitting data over phone lines at 30 characters per second (CPS); (3) by means of a teletype (TTY) connected to the same computer, transmitting data at 10 CPS. All terminals were located in separate air-conditioned rooms to insure privacy. Separate rooms were also used to administer the paper-pencil form of the MMPI. Only the 425 items that are scored on the basic MMPI scales were administered. In the computer conditions, responses were recorded by pressing "T" for a true response, "F" for a false response, and "?" for a "cannot say" response. Responses were then transmitted to the computer by having the testees press the "RETURN" key.

Procedure. Students were randomly assigned to one of the experimental conditions. All testees were given a brief description of the task and were read a set of general instructions, which differed slightly between paper-pencil and computer conditions. Instructions included information on how to record their responses as well as assurances regarding the confidentiality of their answers. Students read the instructions along with the experimenter. Each experimenter administered approximately one-third of the MMPIs in each condition. In addition to the general instructions, students in both computer conditions were given some practice with the computer by having them record their name, student identification number, and sex in the presence of the experimenter. In each condition after the student understood all of the instructions, the experimenter left the room. Students reported to the experimenter upon completing the task.

Analysis. Three types of analyses were employed in order to assess the differences among paper-pencil, CRT, and TTY administrations of the MMPI. At the profile level, differences

Table 1
Means, Standard Deviations, and Analyses of Variance
of MMPI Scales Administered in Different Formats

Scale	Paper-pencil (N=45)		CRT (N=37)		Teletype (N=44)		F-Ratio	p-Value
	\bar{X}	S.D.	\bar{X}	S.D.	\bar{X}	S.D.		
?	1.889	3.002	15.757	18.435	14.864	15.409	13.994	.00000334
L	2.622	2.124	2.730	2.232	2.932	1.835	.257	.774
F	6.113	3.817	6.432	4.616	5.364	3.321	.825	.441
K	14.311	5.134	13.297	5.720	15.136	4.338	1.329	.269
1	5.422	4.218	5.054	3.274	4.614	3.565	.522	.594
2	19.311	5.888	18.243	5.118	18.500	4.112	.505	.605
3	21.000	4.390	18.865	3.772	20.955	4.367	3.292	.040
4	17.622	4.858	15.865	4.923	16.636	4.641	1.382	.255
5	30.311	5.368	28.541	5.714	28.477	4.901	1.669	.193
6	11.067	3.434	9.432	2.930	9.977	3.144	2.836	.062
7	15.400	8.593	14.108	8.714	13.796	6.890	.489	.615
8	16.178	9.623	15.432	9.720	14.023	7.519	.658	.520
9	19.289	4.294	18.487	5.026	18.955	4.803	.297	.744
0	26.400	10.532	28.892	10.661	24.977	9.105	1.530	.221
MANOVA all 14 scales							1.661	.024
MANOVA 13 scales (excluding the ? scale)							.806	.737

among groups were assessed by means of a multivariate analysis of variance (MANOVA) of raw scale score data. Also computed were the univariate analyses of variance (ANOVA) for each scale. At the item level, chi-square analyses were performed on each item in order to determine differences in item endorsements among the three groups. A comparison of two-point code-type frequencies across groups was also made.

Results

Means, standard deviations, and the results of the MANOVA and ANOVAs are shown in Table 1. The MANOVA revealed significant differences ($p < .05$) among the mean profiles of the three groups. When the data were analyzed with the ? scale excluded, these profile differences were no longer found ($p > .70$), suggesting that it was primarily the ? scale which contributed to these differences. Inspection of the univariate ANOVAs for each scale provided evidence to support this. While some of this difference may have come from scales 3 (Hysteria) and 6 (Para-

noia), it is apparent that the F ratio associated with the ? scale is so large that it alone could account for the mean profile differences. Examination of the scale means revealed that the largest difference among the three groups occurred on the ? scale, with each of the computer conditions eliciting an average of about 15 non-responses and the paper-pencil condition averaging about 2 non-responses. The largest variability within groups also occurred on the ? scale for the computer conditions.

Individual items were compared for differential response patterns using two separate sets of chi-square analyses (condition \times response). In the first set, omitted items were coded as missing data, whereas, in the second set, they were coded as valid responses. Since this study was seen as exploratory in nature, the .10 level was considered satisfactory for statistical significance. It was found in the first analysis that 42 items showed significant chi-squares, and in the second analysis that there were 63 significant items. The greater number of significant items when non-responses were included as valid responses may be spuriously high, since many of those

items had expected frequencies as low as one or two. These items do not consistently come from any particular MMPI scales; rather, they distribute across all the scales.

Results of an analysis of two-point codes across groups showed that individuals were distributed across a wide range of profile types both within and between groups; consequently, there were a large number of code types with very low or zero frequencies. Thus, given the sample size of this experiment, information about frequency of occurrence of particular codetypes did not appear to be very useful.

Because the results of this experiment showed large numbers of omitted responses in the computer conditions, a second experiment was designed to determine if: (1) the large number of omitted responses was due to a mode of administration effect in the computer condition; (2) the number of items omitted was a result of the different methods of non-responding (i.e., leaving an item blank in the paper-pencil condition, as opposed to pressing the ? and RETURN keys in the computer condition); or (3) the number of items omitted was due to differences in directions given to testees in the computer conditions.

Experiment 2

Method

Subjects. Thirty-nine male undergraduates in the introductory psychology course at the University of Minnesota took part in the experiment.

Experimenters. Two trained undergraduate psychology majors served as the experimenters.

Materials and apparatus. Because the TTY condition took about twice as long to administer as the CRT condition and had no added testing benefits, only the paper-pencil and CRT conditions were administered. The instructions and answer sheet in the paper-pencil condition were modified to allow the testees to respond “?” to an item they wished to omit rather than having them leave the item blank. The method of responding to omitted items in the CRT condition

was also changed in order to approximate more closely the passive method of omitting an item in the paper-pencil condition (i.e., leaving the item blank). If respondents wanted to omit an item, they had to press only the RETURN key, rather than pressing both the ? and RETURN keys. Directions in the CRT condition were changed to reflect this procedural modification, and also to more closely correspond to the standard directions given in the paper-pencil administration.

Procedure. The procedure was essentially the same as that in Experiment 1, incorporating the changes made in the test formats. In addition, *t* tests were computed comparing means on the ? scale between Experiment 1 and Experiment 2 on both the paper-pencil and CRT administrations.

Results

Table 2 contains the means, standard deviations, and the results of the MANOVA and ANOVAs. An inspection of Table 2 shows there was no significant mean profile difference between the modified paper-pencil and computer administrations of the MMPI. In addition, no significant difference was found between conditions on the ? scale, with each group omitting about 11 items on the average. The only scale showing a significant difference between conditions was scale 6, which showed higher scores for the paper-pencil group than the CRT group, just as in Experiment 1.

Closer examination of the scale 6 means and standard deviations in both experiments reveals little absolute difference between the groups. A significant difference was found on the ? scale for the paper-pencil administrations between the experiments, with a higher mean number of omitted items [$t(64) = 3.22, p < .01$] in Experiment 2. No significant difference was found on the ? scale for the CRT administrations between experiments [$t(53) = 1.28, p > .10$].

In the between-group comparison of responses to individual items, 17 items showed significant chi-squares when omitted items were

Table 2
Means, Standard Deviations, and Analyses of Variance
of MMPI Scales Administered in Different Formats

Scale	Modified Paper-pencil (N=21)		Modified CRT (N=18)		F-Ratio	p-Value
	\bar{X}	S.D.	\bar{X}	S.D.		
?	11.333	12.550	10.278	12.300	.070	.793
L	3.286	2.305	2.778	1.801	.573	.454
F	6.143	3.928	6.722	5.108	.160	.691
K	14.524	5.653	15.167	3.792	.168	.685
1	5.619	4.353	4.444	4.260	.720	.402
2	20.714	4.440	19.833	6.233	.264	.611
3	21.714	4.880	19.556	4.355	2.092	.156
4	17.952	4.455	17.333	4.777	.175	.678
5	29.714	4.971	27.833	4.204	1.596	.214
6	11.381	2.479	9.889	2.166	3.938	.055
7	14.619	6.778	13.444	7.687	.257	.615
8	15.857	7.933	13.278	8.969	.909	.347
9	18.238	4.146	17.889	4.934	.058	.811
0	25.429	8.942	26.889	11.210	.205	.654
MANOVA All 14 Scales					.648	.799

treated as valid responses. Only one of these items (number 302) was an item which produced a significant difference in both experiments; other previously significant items did not replicate across experiments.

Discussion

In the present study, significant differences were found between paper-pencil and computer administrations of the MMPI on both scale 6 and the ? scale. Essentially both types of administrations produced otherwise similar profiles when instructions for both were made nearly equivalent. Differences on scale 6 were small, and the clinical significance of this difference needs to be investigated further.

The question remains: how can we account for differences in the number of omitted items between modes of administration? Since there was no significant difference between the means of the CRT administrations between experi-

ments, it can be concluded that the attempt to make responding to omitted items more passive had failed. The computer condition, either because of the design of the hardware or because of the nature of the software, forced respondents to make an active response to omit an item; testees had to press the ? key and/or the RETURN key in order for the next item to be presented. On the other hand, in the standard paper-pencil administration testees had to make a passive response to omit an item, i.e., leaving the appropriate space on the answer sheet blank. It is reasonable to assume that because an active response had to be made in the computer conditions of both experiments, the number of omitted items would not decrease significantly from the first to the second experiment. In light of the results of the second experiment, it appears that differences on the ? scale were not due to differential reactions to the paper-pencil and computer conditions per se, but rather to the differential response options offered by each method. When the mode of non-response was

made more active in the paper-pencil condition (responding ?, rather than leaving the item blank), the number of non-responses increased significantly.

Although there do not appear to be practical differences on overall profiles between the two modes of administration for most respondents, some way must be found to deal with large numbers of omitted items. Future research might be directed toward two possible approaches. A forced-choice format might be used, to eliminate omitted items. Or the computer might check for large numbers of omitted items and either inform the tester of this fact, or automatically re-administer the entire set of unanswered items with instructions to give the best possible response. In addition, researchers might be interested in determining which characteristics distinguish those testees who omit many items from those who omit few items.

Another consideration is that of the generalizability of the results of this study. Although the computer-administered MMPI seems to be reasonably equivalent to the conventional paper-pencil administration in a "normal" college population, there is no reason to expect necessarily that these results would generalize to a psychiatric population. For example, it might be reasonable to expect that a paranoid individual would not react to the computer in the same way as either a college student or an individual with another psychiatric diagnosis. It is possible that significant interactions may exist between person-types and mode of administration. Future research in this area should probably include samples from different psychiatric and nonpsychiatric populations, and should separately examine responses to computer-administered inventories vs. paper-pencil inventories for each different person-type.

Finally, if computerized administration of personality inventories is shown to be appropriate for a variety of sub-populations, the next goal should be to investigate the computer's ability to administer a unique subset of items to a testee. Research is already underway on the principles of efficiently branching a person

through an item pool in the areas of ability measurement (Weiss, 1975; Weiss & Betz, 1973), personality measurement (Kleinmuntz & McLean, 1968), and diagnostic classification (Biskin, 1975; Butcher, Clavelle, & Hoffman, 1974; Morf & Krane, 1973). Although further research is certainly necessary, preliminary evidence from the present study indicates that the effects of computer administration on the accuracy of obtained personality inventory scores may be minimal, increasing the probability that branching models of item presentation can confidently be introduced in the administration of personality inventories.

References

- Biskin, B. H. *The administration of personality tests by computer: some substantive and methodological issues*. Unpublished manuscript, University of Minnesota, 1975.
- Butcher, J., Clavelle, P., & Hoffman, N. *Clinical limitations of MMPI short forms and a strategy for developing a tailor-made MMPI short form*. Paper presented at the 82nd Annual Convention of the American Psychological Association, New Orleans, September, 1974.
- Dunn, T., Lushene, R., & O'Neill, H. *The complete automation of the Minnesota Multiphasic Personality Inventory and a study of its response latency* (Technical Memo No. 28). Tallahassee: Florida State University, Computer-Assisted Instructional Center, 1971.
- Finney, J. C. *Manual for psychiatrists and psychologists: psychodiagnostic consultation service*. Finney Institute for the Study of Human Behavior. Lexington, Kentucky: Finney Institute for the Study of Human Behavior, 1969.
- Fowler, R. Computer interpretation of personality tests: the automated psychologist. *Comprehensive Psychiatry*, 1967, 6, 455-467.
- Johnson, J., & Williams, T. The use of on-line computer technology in a mental health admitting system. *American Psychologist*, 1975, 30, 388-390.
- Kleinmuntz, B., & McLean, R. Diagnostic interviewing by digital computer. *Behavioral Science*, 1968, 13, 75-80.
- Lushene, R., O'Neill, H., & Dunn, T. *Equivalent validity of a completely computerized MMPI* (Technical Memo No. 48). Tallahassee: Florida State University, Computer-Assisted Instruction Center, 1972.

- Morf, M., & Krane, W. A strategy for developing automated assessment systems. *Behavioral Science*, 1973, 18, 294-306.
- Rome, H., Swenson, W., Mataya, P., McCarthy, C., Pearson, J., & Keating, R. Symposium on automated techniques in personality assessment. *Proceedings of the Mayo Clinic*, 1962, 37, 61-82.
- Smith, R. Examination by computer. *Behavioral Science*, 1963, 8, 76-79.
- Weiss, D. J. (Ed.). *Computerized adaptive trait measurement: problems and prospects* (Research Report 75-5). Minneapolis: University of Minnesota, Department of Psychology, November, 1975. (NTIS No. AD-018675)
- Weiss, D. J., & Betz, N. E. *Ability measurement: conventional or adaptive?* (Research Report 73-1). Minneapolis: University of Minnesota, Department of Psychology, 1973. (NTIS No. AD-757788)
- Wiggins, J. S. Content dimensions of the MMPI. In J. N. Butcher (Ed.), *MMPI: research developments and clinical applications*. New York: McGraw-Hill, 1969.

Acknowledgements

The authors thank James Butcher and David Weiss for their support in the early stages of this research. We also appreciate the aid of Gary Kinser in the development of several computer programs. In addition, we acknowledge the help of William Iacono in the initial phase of the study.

Author's Address

Bruce Biskin, Department of Psychology, University of Maryland, College Park, MD 20742