

Equivalence of Conventional and Computer Presentation of Speed Tests

Valerie A. Greaud and Bert F. Green
The Johns Hopkins University

This study examined the effects of computer presentation on speeded clerical tests. Two ratio scores—average number of correct responses per minute and its inverse, average number of seconds per correct response—were examined as variants of the conventional score, number of correct responses in a fixed interval of time. Ratio scores were more reliable than number-correct scores and were less sensitive to testing time. Tests administered on the computer were found to be at least as reliable as conventionally administered tests, but examinees were much faster in the computer mode. Correlations between paper-and-pencil and computer modes were high, except when task differences were introduced by computer implementation.

When a conventional paper-and-pencil test is transferred to a computer for administration and scoring, there is no assurance that the scores achieved with computer presentation will be comparable to those obtained with the conventional format. Even though the content of the items is the same, mode of presentation could make a difference in test-related behaviors, such as the propensity to guess, the facility with which earlier items can be reconsidered, and the ease and speed of responding.

Response speed is particularly an issue on highly speeded clerical and perceptual tests. The notable characteristic of such tests is that the requisite task,

such as crossing out Os, counting Xs, or making simple perceptual comparisons, is extremely easy. Examinees differ, however, in their ability to perform the task quickly. It is not very interesting to consider whether or not examinees can perform the task because, with unlimited time, most are able to select the correct answers. For these tasks, it is more interesting to assess how quickly examinees respond, because reliable and valid individual differences in response speed do exist. Other types of speeded tests require more complex skills or abilities, such as verbal or mathematical reasoning, and there is a limit on the amount of time available to answer the items. On such speeded tests, examinees usually differ in skill and knowledge as well as in speed. To avoid confusion with these terms, the simple speeded tests are referred to here as “clerical tests”.

If the same test is given in paper-and-pencil form and on a computer, the similarity of scores in the two modes depends on the ability measured and the type of score used. Most studies of the mode difference have involved tests that are more complex than clerical tests. For example, Vinsonhaler, Molineaux, and Rogers (1968) compared untimed accuracy scores on tests of verbal ability. They found no difference between examinees' scores on computer-presented tests and conventional tests, with the minor exception of the slightly higher scores achieved by the very low ability examinees on computer-administered tests. They found no dif-

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 10, No. 1, March 1986, pp. 23–34
© Copyright 1986 Applied Psychological Measurement Inc.
0146-6216/86/010023-12\$1.85

ference between modes in the reliability of scores. Vinsonhaler et al. also recorded times to complete the tests and found that examinees were slower on the computer. However, their computer system had printed output, which, as Vinsonhaler suggested, is inefficient.

Sacher and Fletcher (1978) found no score difference between modes of administration for a verbal reasoning task, but found higher scores with paper-and-pencil administration of a symbolic reasoning task. Both tests had time limits, but the length of the limits with respect to the number of items on each test affected the tests differently, as Weiss (1978) pointed out. The verbal reasoning test was essentially untimed; most examinees reached the end of the test within the time limit. Performance on the symbolic reasoning test was affected more by the time limit; very few examinees reached the end of this test.

Of greater relevance to the present study, Lansman, Donaldson, Hunt, and Yantis (1982) compared performance on paper-and-pencil and computerized versions of three tests: letter matching, sentence verification, and mental rotations. They found that mean reaction time for correct answers on each of the computerized versions was faster than mean reaction time per correct item on the corresponding paper-and-pencil versions. This difference could occur as a result of faster responding on the computer, or it could be due to the slight difference in the measures. On the computer, reaction times were averaged over correct responses only. On the paper-and-pencil versions, the mean was computed by dividing the time limit for a test by the number of correct responses. Thus, the time taken by incorrect responses was included in the paper-and-pencil mean reaction time measure but not in the computer measure. Percentage of errors was also computed on each version of the tests. The percentage of errors on the letter-matching test was lower for the paper-and-pencil version than for the computerized version, suggesting a trade-off between speed and accuracy. This measure was not reported for the other tests.

The difference between presentation modes is a practical problem for the clerical tests, Numerical Operations (NO) and Coding Speed (CS), of the

Armed Services Vocational Aptitude Battery (ASVAB). The ASVAB is now being adapted for computer presentation. Eventually this battery will be presented on the computer only, but for an interim period, both computer and paper-and-pencil versions will be in use. It is important that the computer version of the battery be as similar as possible to the paper-and-pencil version.

The ASVAB contains 10 subtests. Eight of the subtests are tests of knowledge, skill, or ability, and computerized adaptive versions of these are being developed. The other two, NO and CS, are highly speeded clerical tests and will be administered nonadaptively by computer. Because response time is a critical component of an examinee's score on clerical tests, anything that affects response time is likely to affect test performance. For example, on a paper-and-pencil version of the ASVAB, Wegner and Ree (1985) found that response time was slowed significantly if the response blanks on the answer sheet were arranged differently from the items in the exam booklet. This finding demonstrates the sensitivity of clerical tests to response method modifications.

In a preliminary comparison of computer and paper-and-pencil versions of the NO and CS tests, paper-and-pencil forms were transferred to the computer. Initial trials showed that responses were made much more quickly on the computer than with paper-and-pencil (Vicino & Hardwicke, 1984), and many of the experimental examinee group were able to complete all items in the computer-presented tests within the paper-and-pencil time limits. To remedy this difference in response speed, the time limits for the computer-presented versions were shortened.

An obvious difference between paper-and-pencil and computerized presentation modes is the method of responding. On paper-and-pencil tests, examinees usually respond by locating and filling in a bubble on an answer sheet; on computer-presented tests, examinees must press a response key on a computer terminal. Pressing a computer terminal key rather than marking a bubble on an answer sheet is likely to affect response time. Rather than simply shortening the time limits for the computer versions of the two clerical tests, it seems wise to

consider a variety of scoring methods and to make use of the computer's ability to keep track of elapsed time.

This issue will arise whenever a highly speeded clerical test is given by computer, so it is pertinent to examine the properties of some alternative scores and to consider the implications for establishing score scale equivalence with the conventional version of the test. An important practical consideration with the ASVAB is that, for a considerable period of time, both computerized and conventional versions must exist simultaneously. The scores on the two versions must be calibrated so that an examinee will obtain an equivalent score no matter which version he or she takes. The need for calibration occurs whenever a test is modified for computer administration and, as many more tests are computerized, such calibration will become more common.

Because of the limitations of group testing, conventional clerical tests are designed to contain more items than most examinees can answer in a fixed testing interval. An examinee's score is the number of items answered correctly in this interval of time. The time limit must be carefully chosen; if it is too long, many examinees will complete the test and their scores will be essentially indistinguishable—a "ceiling effect". The only differences in their scores will be due to a very few incorrect responses. Of course, if the time limit is set too short, the test is essentially shortened, leading to lower reliability.

Computer administration can avoid the problem of selecting the correct time limit by presenting a fixed number of items and recording the time taken by each examinee to answer the entire set of items. This method of test administration also permits scoring of examinees' performance with either of two ratio measures: (1) number of correct responses per unit time or (2) its inverse, time per correct response. Conventional number-correct scoring with fixed testing time limits would give the same score to any examinees correctly finishing all items before the time limit. A ratio score differentiates among these examinees according to their actual speed of completion.

Other practical issues to be considered when a

test is modified are reliability and comparability of scores in each version. The original test serves as the standard by which the modified version is judged. The reliability of scores in the newer version should be at least as great as that of scores in the original. Equivalence of scores in the two versions must also be established to provide evidence that both versions are measuring individuals in a similar manner.

The present study examined four questions.

1. Which measure of speeded performance is best? The current measure, number of correct responses in a fixed time interval, was compared with two ratio measures, a measure of response rate and a measure of reaction time.
2. Do computer and paper-and-pencil presentation modes differ in reliability? It had been expected that reliability might be greater for computerized tests than for paper-and-pencil tests because administration procedures on the computer could be less variable than conventional procedures.
3. How much do mean scores differ between modes? Based on the studies mentioned above, examinees could be expected to work considerably faster on the computer-presented tests.
4. How comparable are scores between modes? The equivalence between computerized tests and paper-and-pencil tests was assessed by correlating scores between modes.

Method

Materials

Shortened versions of NO and CS, the two speeded clerical tests of the ASVAB, were used in this study. Items from retired ASVAB forms were selected to create the shortened tests. For the paper-and-pencil tests, the problems were presented on one sheet of paper and examinees made their responses on a separate bubble-type answer sheet. Individuals' times to complete each test were recorded and the answer sheets were scored by hand. The computerized forms were presented on a TVI-920 computer terminal. Examinees indicated their responses by pressing one of the labeled keys arranged side-by-side in the top row of the keyboard. The master program

recorded response times for each item and scored all responses.

The Numerical Operations test forms in this experiment each consisted of 25 simple multiple-choice math problems. For each problem, examinees were asked to select the correct answer from among four options. An example of a NO problem is:

- 2 + 3 = _____
 A. 1
 B. 5
 C. 6
 D. 10

The paper-and-pencil version presented the 25 problems in four columns on a single sheet of paper. On the computer, each problem was presented individually in the center of the screen.

The Coding Speed test forms consisted of 28 multiple-choice coding problems. A table containing 10 words and their corresponding four-digit codes was presented at the top of the test page or the computer screen. Each problem contained one word and five codes. The examinees' task was to

select the code that corresponded to the word, according to the conversion table. For the paper-and-pencil version, all 28 problems were presented on one sheet of paper in four groups of 7 problems (see Figure 1). On the terminal, each problem appeared individually in the center of the computer screen.

Procedure

Fifty college students each took two forms of the NO test by paper-and-pencil and two forms on a computer terminal. They also took two forms of CS by paper-and-pencil and two on the computer terminal, for a total of eight tests per student. All of the students were individually tested; they were encouraged to work quickly but were given as much time as they needed to complete each test.

Testing order was counterbalanced by mode (computer or paper-and-pencil), by test type (NO or CS), and by form. To achieve this balanced design, examinees were assigned to one of two groups. The 25 examinees in Group 1 took the four paper-and-pencil test forms first and the four computer-

Figure 1
 Example of Coding Speed Item Arrangement
 on Paper-and-Pencil Forms

bargain....8385	house.....2859	owner.....6227			
chin.....8930	knife.....7150	point.....4703			
game.....6456	music.....1117	sofa.....9645			
	sunshine...7489				
	A	B	C	D	E
1. game	6456	7150	8385	8930	9645
2. knife	1117	6456	7150	7489	8385
3. bargain	2859	6227	7489	8385	9645
4. chin	2859	4703	8385	8930	9645
5. house	1117	2859	6227	7150	7489
6. sofa	7150	7489	8385	8930	9645
7. owner	4703	6227	6456	7150	8930
	A	B	C	D	E
8. music	1117	2859	7489	8385	9645
9. knife	6227	6456	7150	7489	8485
10. sunshine	4703	6227	6456	7489	8930
11. chin	1117	2859	4703	7150	8930
12. sofa	4703	6227	7150	8485	9645
13. bargain	2859	6456	8385	8930	9645
14. point	1117	4703	6227	6456	7150

ized test forms second, beginning with a NO test form and alternately taking NO and CS tests. The 25 examinees in Group 2 took the same test forms as Group 1, but in reverse order, beginning with a CS test form and taking the computerized forms before the conventional forms.

On the computer-presented tests, examinees occasionally held the response option key down too long or pressed the RETURN key after the response key. These faulty responses, which comprised less than 2% of responses, were deleted from the data.

Results and Discussion

As mentioned earlier, the score usually obtained for paper-and-pencil tests is the number of correct responses in a fixed time period (NC). Recall that, in this experiment, the examinees were not tested for a fixed interval; they completed all items and their individual times were recorded. The responses collected from computerized presentation were used in order to examine the effects of selecting different time limits for NC. Because the computer supplies individual item response times, various testing time limits for the same response data could be simulated by imposing different time limits after the fact.

Time limits were selected in relation to examinees' computerized test completion times. Five time limits (the mean test completion time and one and two standard deviations above and below the mean) were selected, yielding time limits for NO of 94, 81, 69, 56, and 43 seconds. For CS, the time limits were 150, 132, 114, 97, and 79 seconds.

Alternate-forms reliability coefficients of NC scores for each time limit are presented in the first column of Table 1. For both NO and CS, the correlation between forms decreased as the time limits increased. Longer time limits allowed more examinees to complete all items; for examinees who completed all items, the NC scores differentiated between them only on the basis of their very few incorrect responses.

Two ratio scores, CPM (average number of correct responses per minute) and SPC (average number of seconds per correct response), were also calculated for the computerized test forms based

on the time limits used above. Alternate-forms reliabilities of CPM and SPC scores for each time limit are also presented in Table 1. These reliabilities vary slightly with the time limits, but not in the manner nor to the extent that correlations between NC scores do. Note that, for all scores, correlations decreased with the shortest time limits; of course, a test composed of few items has low reliability.

The reliability of NC scores is clearly very sensitive to testing time limits. Longer time limits increase the ceiling effect and reduce reliability. The powerful effect of time limit selection on reliability demonstrates that number-correct scoring is of limited value. The appropriate time limit is difficult to determine and it depends on the performance of the group being tested. Ratio measures should be better because they are less dependent on the length of the testing interval. The ratio scores are affected less by time limits than NC scores because individual completion times are used in score computation. Examinees who finish before the time limit can be credited for working quickly. Number-correct scores cannot credit examinees for finishing early.

Descriptive statistics for scores based on unlimited and limited testing times are shown in Table 2. The time limits chosen were those that yielded the most reliable NC scores (see Table 1). For clarity, data are only shown for one of the computerized

Table 1
 Within-Mode Correlation Coefficients
 Between Computerized Test Forms
 for NC, CPM, and SPC

Test and Time Limit	NC	CPM	SPC
Numerical Operations			
94 seconds	.55	.79	.74
81 seconds	.48	.79	.73
69 seconds	.63	.77	.69
56 seconds	.74	.78	.72
43 seconds	.69	.68	.61
Coding Speed			
150 seconds	.54	.74	.77
132 seconds	.51	.73	.76
114 seconds	.60	.73	.76
97 seconds	.66	.70	.73
79 seconds	.62	.62	.63

Table 2
Descriptive Statistics for NC, CPM, and SPC
on Computer Presented Tests

Test and Statistic	NC	CPM	SPC
Numerical Operations			
Unlimited testing time			
mean	23.76	22.44	2.76
standard deviation	1.69	3.95	.53
skewness	-3.21	.01	1.01
Limited testing time (56 seconds)			
mean	20.38	22.59	2.74
standard deviation	3.03	3.86	.53
skewness	-.87	-.10	1.40
Coding Speed			
Unlimited testing time			
mean	27.40	15.19	4.02
standard deviation	1.23	2.10	.60
skewness	-2.96	-.20	.88
Limited testing time (97 seconds)			
mean	23.80	15.10	4.05
standard deviation	3.02	2.12	.60
skewness	-.50	-.09	.78

forms (Form 3) of NO and of CS; data for the other forms are very similar. The descriptive statistics indicate that NC scores based on a limited testing time differ from NC scores based on unlimited testing time. Reducing the testing time lowers the mean, increases the standard deviation, and reduces the skewness of the distribution of NC scores. Note that these differences do not occur for CPM and SPC.

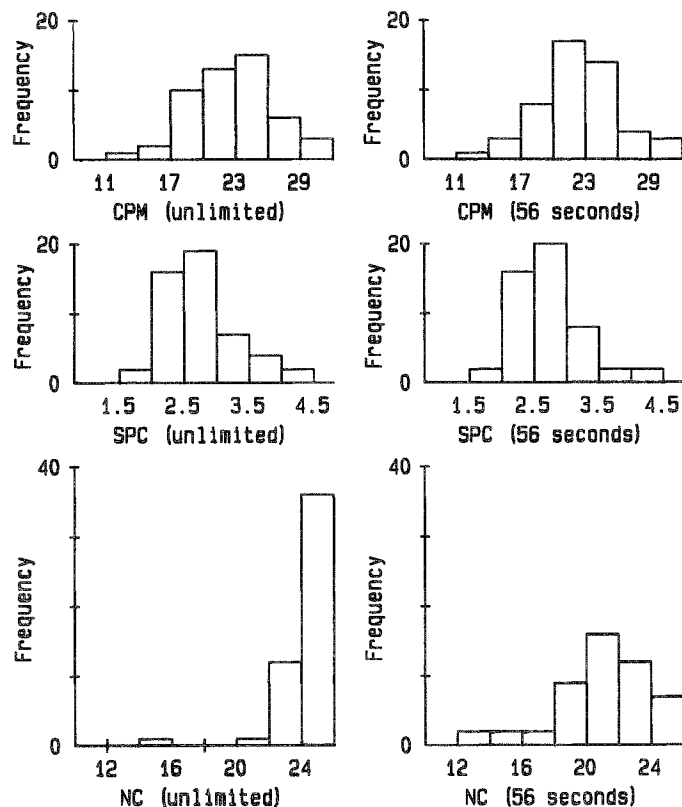
Figures 2 and 3 show grouped frequency distributions of scores based on unlimited and limited testing times. Of course, the NC score distributions are extremely skewed if testing time is unlimited. For the chosen time limits, the distributions of NC scores are more reasonable, but still quite skewed. The CPM and SPC distributions scarcely differ with changes in the testing time.

CPM and SPC produce exactly reverse ordering of scores because one is the inverse of the other, specifically, $60 \times \text{CPM} = 1/\text{SPC}$. However, because they are not linearly related, their score distributions differ and one scale may be preferable. Comparison of SPC and CPM indicates that SPC distributions are more skewed than CPM (see Table 2,

Figure 2, and Figure 3). As experimental psychologists know well, reaction time distributions generally are skewed. When doing analyses of variance on reaction time data, it is desirable to use a transformation. Use of CPM, the reciprocal of SPC, is advocated. Added advantages are that CPM has a true zero, and that higher CPM scores correspond with better performance. The remaining analyses will, therefore, be presented only for CPM.

Conventional time-limited number-correct scores explicitly penalize errors. A similar error penalty was incorporated in the ratio scores by counting only correct responses, but timing both correct and incorrect responses. With tests that encourage quick responding, there is a possibility that accuracy will be sacrificed for speed. That is, perhaps faster examinees make more errors. Table 3 presents average correlation coefficients between number of responses per minute (counting all responses, not only correct ones) and number of errors. The one positive correlation is quite small and it is not significant. This indicates that examinees do not make more errors while attempting to respond quickly.

Figure 2
 Histograms of CPM, SPC, and NC Scores Based on Unlimited
 and Limited (56 seconds) Testing Time for Numerical Operations



Average number of errors is also shown in Table 3, demonstrating how few errors were made on these tests.

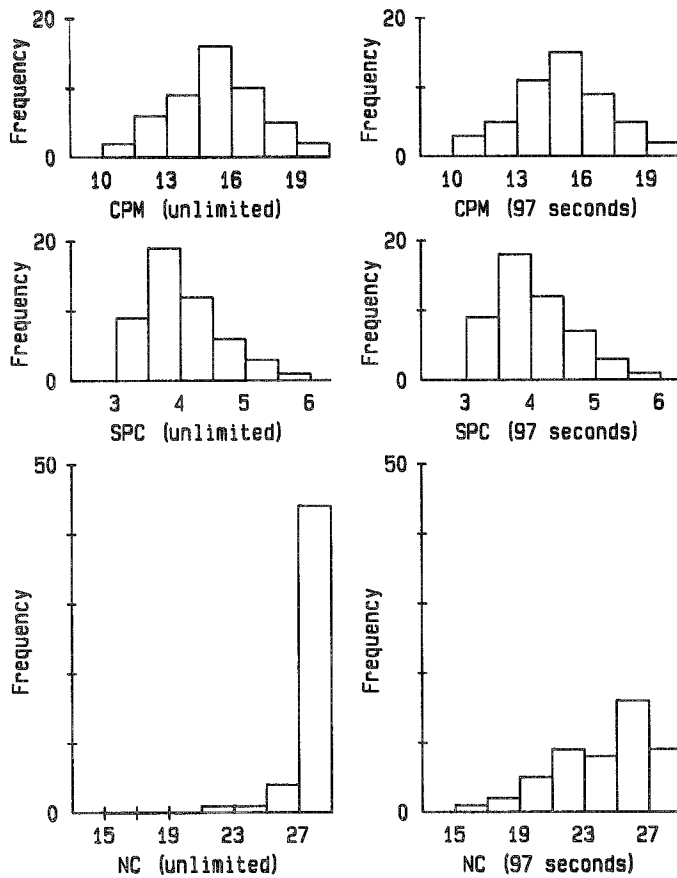
Reliabilities of scores in each mode were assessed by correlating scores on alternate forms. Correlation matrices for each test are presented in Table 4. For each test, Forms 1 and 2 were administered by paper-and-pencil and Forms 3 and 4 were administered on the computer. The NO alternate-forms reliability of .77 for the computerized forms is significantly greater ($p = .05$) than .62 for the paper-and-pencil forms, by Steiger's (1980) correction for dependent correlations. There is no significant difference between the CS alternate-forms reliability coefficients of .74 on the computer and .70 with paper-and-pencil. These comparisons show

that, for the tests used in this experiment, the reliability of computer presentation equals or exceeds that of paper-and-pencil presentation.

Graphs of the mean CPM scores and their standard deviations for each group of examinees on each form are shown in Figure 4. The mean scores for NO are 15.63 on the paper-and-pencil forms and 21.45 on the computerized forms. The respective means for CS are 9.15 and 14.82. Scores increased from paper-and-pencil administration to computerized administration by 37% for the NO tests and by 62% for the CS tests.

The analyses of variance for both tests show significant differences ($p < .01$) for Mode, Form within Mode, and Group by Form within Mode (Table 5). The most prominent difference in scores

Figure 3
 Histograms of CPM, SPC, and NC Scores Based on Unlimited
 and Limited (97 seconds) Testing Time for Coding Speed



is due to Mode. Examinees are very much faster at taking these speeded tests on the computer than with paper-and-pencil. The significant Form within Mode effect means that the average scores within presentation modes differed. The observed slight mean difference in response time for test forms within a mode occurred because the forms were not completely balanced for type of problem. For example, on NO, response times are longer for division problems than for the other problem types. On CS, response times are longer when the correct alternative is further to the right among the five options (see Figure 1). The experimental difference between groups is order of test administration,

so the Group by Form within Mode effect is an interaction of differences due to forms and the order of presentation of these forms. This effect appears to be primarily the result of a reversal in group scores on one form of each test (see Figure 4).

The significant Mode effect might be interpreted as being due simply to a difference between forms, because Forms 1 and 2 were always administered with paper-and-pencil and Forms 3 and 4 were always administered on the computer. There are two reasons why this hypothesis is unlikely. First, the size of the difference between modes is much larger than that between forms. Estimated com-

Table 3
 Average Number of Errors and Average Correlations
 Between Number of Responses Per Minute and
 Number of Errors

Test and Mode	Number of Errors	Correlation
Numerical Operations		
Paper-and-pencil	1.04	- .18
Computer	1.31	- .26
Coding Speed		
Paper-and-pencil	.92	- .11
Computer	.39	.11

ponents of variance for NO are 16.43 for Mode and .93 for Form within Mode. The respective variance components for CS are 15.96 and .15. Second, the actual ASVAB subtest forms were equivalent. For each type of test used in this experiment, the two forms presented in each mode were created by dividing a subtest from an actual retired form of the ASVAB. For example, NO Forms 1 and 2 were originally two halves of one form of a NO subtest, and Forms 3 and 4 were two halves of another. These whole subtests have been shown to be parallel (Rec, Mullins, Matheus, & Massey, 1982). So the average score on Forms 1 and 2 would be expected to be the same as the average on Forms 3 and 4. The finding that they are different can safely be attributed to a mode difference.

A mean difference between modes can easily be adjusted in scoring. However, in order for the corrected computer scores to be considered equivalent to the paper-and-pencil scores, the two sets of scores must correlate highly. The data regarding correlations between modes differ for the two tests used in this experiment, so the tests will be discussed separately.

For NO, the between-mode correlations range from .56 to .68 (Table 4). To determine what the between-mode correlations would be if the tests had been perfectly reliable, the classical correction for attenuation can also be computed. When corrected for attenuation, the correlations range from .81 to .98 (Table 6). These values are high, but they are not perfect. Some small difference ap-

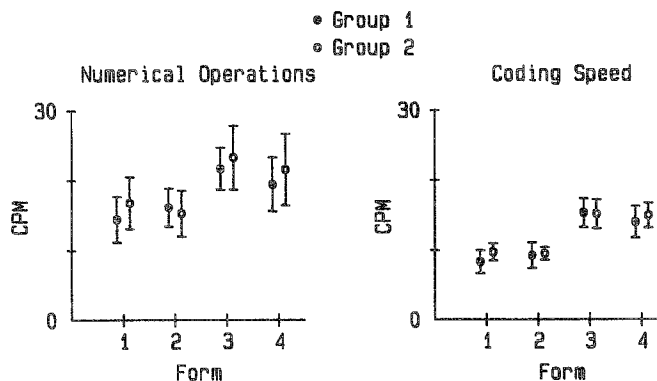
parently was introduced by the change in response method. The effect should be smaller if the tests are of regulation length and probably can be ignored in practice, but further study is indicated.

For CS, however, the between-mode correlations were quite low, ranging from .28 to .61 (Table 4); and they remained low, .38 to .84, when corrected for attenuation (Table 6). This indicates that the task was different in the two modes. On the computer, each problem was presented individually in the center of the computer screen, whereas the items were printed in blocks of seven in the conventional test booklet. Apparently, keeping track of the location of successive problems within each block of items is part of the task, because this simple change had a strong effect on test performance. If the items had been presented in blocks of seven on the computer, the tasks in the two modes might

Table 4
 CPM Within-Mode and Between-Mode
 Reliability Coefficients for
 NO (Lower Triangle) and
 CS (Upper Triangle)

Form	Form			
	1	2	3	4
1	---	.70	.28	.39
2	.62	---	.53	.61
3	.56	.56	---	.74
4	.67	.68	.77	---

Figure 4
 CPM Means and One-Standard Deviation Intervals
 for Each Group of Subjects on Each Form



have been more similar and the between-mode correlations probably would have been higher.

The implication of this task difference is that care must be exercised in transferring a test from paper-and-pencil to the computer. Modifications in the way items are presented can affect performance, especially on tests such as CS. Although both NO and CS are speeded clerical tests, the req-

uisite abilities for these tests are slightly different. The NO test requires recognition of numerical products; performance is not strongly affected if items are presented individually rather than in a group. CS, however, requires matching of words and digit codes. On this test, it seems much easier to keep track of the digit code options when items are presented alone.

Table 5
 Analyses of Variance for CPM

Test and Source of Variance	df	SS	MS	F
Numerical Operations				
Group	1	83.32	83.32	1.93
Ss (Group)	48	2067.16	43.07	
Mode	1	1692.83	1692.83	33.91 *
Group x Mode	1	15.26	15.26	.41
Ss x Mode (Group)	48	354.76	7.39	
Form (Mode)	2	100.71	50.36	12.87 *
Group x Form (Mode)	2	67.78	33.89	8.66 *
Ss x Form (Group x Mode)	96	375.68	3.91	
Coding Speed				
Group	1	18.67	18.67	2.17
Ss (Group)	48	412.73	8.60	
Mode	1	1605.94	1605.94	163.94 *
Group x Mode	1	2.12	2.12	.21
Ss x Mode (Group)	48	144.63	3.01	
Form (Mode)	2	16.77	8.39	10.46 *
Group x Form (Mode)	2	16.10	8.05	10.04 *
Ss x Form (Group x Mode)	96	76.97	.80	

* p < .01

Table 6.
 CPM Within-Mode and Between-Mode
 Reliability Coefficients,
 Corrected for Attenuation
 for NO (Lower Triangle)
 and CS (Upper Triangle)

		Form			
Form	1	2	3	4	
1	---	1.00	.38	.54	
2	1.00	---	.74	.84	
3	.81	.81	---	1.00	
4	.97	.98	1.00	---	

Note that the NO tests used here were half the length of NO tests used in practice; the CS tests were one-third the length. Length-corrected reliabilities for NC scores with an optimal time limit from the computer versions were .85 for both NO and CS. These values are slightly higher than those obtained in paper-and-pencil full-length versions of these tests (Stephenson, Matheus, Welsh, Ree, & Earles, in press).

Conclusions

This study has shown, first, that the measure of response rate, CPM, provides a useful scale of performance on clerical tests. It is superior to its inverse, SPC, and to the conventional number correct measure, NC. SPC is a reaction time measure and, as such, produces a skewed distribution of scores. The SPC scale does not have a true zero and scores are inversely related to performance. The distribution and reliability of scores on the NC scale vary with the length of the testing interval. The time limit must be chosen with great care so that distributions will not be skewed and reliability will not be restricted.

Second, it was found that reliability of scores on computer-presented clerical tests was at least as high as for paper-and-pencil tests. This verifies that measurements obtained by the newer computerized presentation are at least as consistent as those obtained in the conventional manner.

The third finding was that examinees responded more quickly on computer-presented clerical tests than on paper-and-pencil versions. That is, examinees are faster at pressing a button than at locating and marking a bubble on an answer sheet. For most cognitive ability tests, the time to indicate an answer is an inconsequential factor because most of the time is spent deciding which answer to choose. But in a clerical test, the tasks themselves are simple, and time to respond is a critical component of examinees' scores.

Fourth, the correlation between modes was high for NO and might have been satisfactory for CS if not for the unintended task difference. Since there were mean score differences between computer and paper-and-pencil modes, either separate norms must be developed for computer-presented tests, or one set of scores must be transformed to achieve equivalence between modes. This is necessary in order to evaluate examinees in comparison with others who have taken the test under different conditions.

Because the computer can record individual item times, an even finer analysis of performance than has been discussed here is feasible. Further research should be done to evaluate the additional information that can be obtained from item responses.

References

Lansman, M., Donaldson, G., Hunt, E., & Yantis, S. (1982). Ability factors and cognitive processes. *Intelligence*, 6, 347-386.

Ree, M. J., Mullins, C. J., Matheus, J. J., & Massey, R. H. (1982). *ASVAB: Item and factor analyses of forms 8, 9, and 10*. Brooks Air Force Base TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Sacher, J., & Fletcher, J. D. (1978). Administering paper-and-pencil tests by computer, or the medium is not always the message. In D. J. Weiss (Ed.), *Proceedings of the 1977 Computerized Adaptive Testing Conference*. Minneapolis MN: University of Minnesota, Department of Psychology, Psychometric Methods Program.

Steiger, J. H. (1980). Tests for comparing elements in a correlation matrix. *Psychological Bulletin*, 87, 245-257.

Stephenson, S. D., Matheus, J. J., Welsh, J. R., Ree, M. J., & Earles, J. A. (in press). *Technical supple-*

- ment to the Counselor's Manual for the ASVAB Form 14. Brooks Air Force Base TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.
- Vicino, F. L., & Hardwicke, S. B. (1984, March). *An evaluation of the utility of large scale computerized adaptive testing*. Paper presented at American Educational Research Association Convention, Chicago IL.
- Vinsonhaler, J. F., Molineaux, J. E., & Rogers, B. G. (1968). An experimental study of computer-aided testing. In H. H. Harman, C. E. Helm, & D. E. Loye (Eds.), *Computer-Assisted Testing: Proceedings of the 1966 CAT Conference*. Princeton NJ: Educational Testing Service.
- Wegner, T. G., & Ree, M. J. (1985). *The 1980 youth population: Correcting the speeded subtests*. Brooks Air Force Base TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.
- Weiss, D. J. (1978). Discussion. In D. J. Weiss (Ed.),

Proceedings of the 1977 Computerized Adaptive Testing Conference (pp. 420–422). Minneapolis MN: University of Minnesota, Department of Psychology, Psychometric Methods Program.

Acknowledgments

This research was supported in part by the Office of Naval Research and the Navy Personnel Research and Development Center through the Personnel and Training Research Programs, Psychological Services Division, Office of Naval Research under Contract No. N 0014-80-K-304.

Author's Address

Send requests for reprints or further information to Valerie A. Greaud, Department of Psychology, The Johns Hopkins University, Baltimore MD 21218.