Journal of Educational Psychology
2001, Vol. 93, No. 4, 772-775

0022-0663/01/$5.00 DOI: 10.1037//0022-0663.93.4.772

Answering Quasi-Randomized Reading Items Without the Passages on the SAT–I

Stuart Katz, Richard L. Marsh, Christopher Johnson, and Erika Pohl
University of Georgia

Examinees can correctly answer many Scholastic Assessment Test (SAT) reading items when the passages accompanying the items are missing. According to one hypothesis, examinees use information from other reading items (*cognates*) belonging to the same passage. The purpose of this study was to test that hypothesis for the revised SAT (SAT–I) reading task. Cognate information was diminished by dispersing cognates among other reading items from 4 parallel forms of the SAT–I. Examinee performance in this condition was inferior to performance in a control, where cognate information was readily available. Nonetheless, the correlation between performance and verbal SAT score in the experimental condition remained high and equivalent to the control. Examinees do seem to use cognate information when passages are missing on the SAT–I reading task, but such information is not among the no-passage factors that systematically affect reading performance.

Examinees can answer many Scholastic Assessment Test (SAT) multiple-choice reading items without the passages that usually accompany these items. Moreover, performance on such a task correlates substantially with verbal SAT score (Katz, Johnson, & Pohl, 1999; Katz & Lautenschlager, 1994; Katz, Lautenschlager, Blackburn, & Harris, 1990). The proffered explanation for these findings is that performance on the task is systematically influenced by factors having nothing to do with the passages themselves, such as test-taking skills and outside knowledge. Among the test-taking skills may be one in which the examinee obtains relevant information from other test items (*cognates*) belonging to the same passage (Powers & Leung, 1995; Pyrczak, 1972; Tuinman, 1973–74). In a study that used the old (pre-1994) version of the SAT, however, no evidence of such a skill was found. Performance without the passages revealed no decline when cognates were dispersed among reading items from other passages (i.e., quasi-randomized), a procedure that effectively eliminated shared or complementary information among the cognates (Katz, Blackburn, & Lautenschlager, 1991). There was, moreover, no decline in the correlation between performance and verbal SAT score. The authors concluded that examinees do not use cognate information on the reading task. Nonetheless, according to Powers and Leung (1995), examinees reported using cognates when answering reading items without the passages on the SAT–I, the revised version of the SAT (*8 Real SATs,* 1996). Although the finding is suggestive, verbal reports are post hoc self-appraisals and do not necessarily reveal what examinees actually do on the reading task or whether what they do actually affects performance. Still, given the changes to the reading task—chief among them fewer but longer passages, and many more cognates per passage—these self-reports do suggest that the role of cognates should be revisited.

Stuart Katz, Richard L. Marsh, Christopher Johnson, and Erika Pohl, Department of Psychology, University of Georgia.

Correspondence concerning this article should be addressed to Stuart Katz, Department of Psychology, University of Georgia, Room 522, Athens, Georgia 30602. Electronic mail may be sent to skatz@arches.uga.edu.

The purpose of this study was to determine whether cognate information actually influences performance on the SAT–I reading task. Using a design similar to that of Katz et al. (1991), reading items from four actual SAT–I tests were administered without the passages. In the control condition, cognates in each passage group appeared together, just as they do on actual tests. In the experimental condition, these same cognates were quasi-randomized across the four tests and two testing sessions.

If shared information among cognates on the reading task contributes to performance when the cognates appear together, performance in the experimental condition should be inferior to that in the control. The reason: Cognate information in the experimental condition, by virtue of being quasi-randomized, should be more difficult to use. Moreover, the correlation between performance and verbal SAT score in the experimental condition should be lower than that in the control, assuming that examinees who perform better on the reading task, with or without the passages, do so at least in part because they use cognate information more effectively. When that information is less usable, the systematic relationship between performance on the reading task and verbal SAT score should diminish. Note that verbal SAT score serves here as a proxy for performance on the reading task with the passages, because reading scores alone are not available. Nonetheless, the proxy should suffice for two reasons. First, the tasks constituting the verbal section of the test are strongly intercorrelated (*ATP Guide,* 1985), and second, the reading task is by far the largest on the verbal section of the SAT–I, comprising more than half of all verbal items.

Method

Participants

Examinees (N = 190) were from an introductory course research pool. Of these, 83 were freshmen and 64 sophomores. Examinees were randomly assigned, in groups of approximately 30, to one of two conditions, control (n = 83) or quasi-random (n = 107). An additional group was assigned to the quasi-random condition to increase sample size (there is limited prior

data for the quasi-random condition, whereas the control has been replicated many times). Both conditions are described in the *Procedure* section.

Examinee verbal SAT scores were obtained from Admissions and Records at the University of Georgia. Among examinees in this study for whom verbal SAT scores were available (n = 173, or 91% of the total sample), verbal SAT scores were recentered (if exams taken were taken in April 1995 or later) or not recentered (if exams were taken prior to April 1995). Recentered scores are those based on a high school graduate reference group more recent than the original 1941 reference group (Feinberg, 1994–95). Mean recentered verbal SAT scores for examinees in the control and quasi-random conditions of this study were 576 (SD = 63, n = 56) and 583 (SD = 65, n = 74), respectively, and these did not differ statistically, $t(128) < 1.0, p > .05$. Mean nonrecentered verbal SAT scores in the control and quasi-random conditions of this study were 507 (SD = 91, n = 21) and 465 (SD = 71, n = 22), and these, too, did not differ, $t(41) = 1.70, p > .05$. (Note that standard deviations for these samples are smaller compared both with national samples, where standard deviations approximate 100, and with those reported by Katz et al., 1990.)

Materials

Reading items were taken from four actual tests administered in 1994–1996 (*8 Real SATs,* 1996). There are 40 items per test, divided among four passage groups. Therefore, total examination length in the experiment was 160 items (16 passage groups), a 60% increase over the number used in Katz et al. (1991). Each passage group contained from 5 to 13 items.

Procedure

Control condition. Examinees were given reading items with the passages deleted and with cognates grouped together on one or two pages, precisely as they appeared on the actual tests. Two of the four tests were administered in a single session, and the remaining two in a subsequent session 2 days later. The tests were given in one ABCD order and its reverse (DCBA). During each session examinees were allowed 30 min per test with a 5-min break between tests.

Quasi-random condition. The same items used in the control condition were used here. However, the items were administered in a single quasi-random order to eliminate access to information provided by cognates. The order was generated as follows:

1. An item was randomly selected without replacement from each of four passage groups in each of the four tests—ABCD, in that order. When all passage groups had been sampled, the process was repeated until all 160 items had been chosen. The resulting list was then partitioned into four successive 40-item groups, or pseudoexams.

2. Wherever necessary, items were redistributed to meet the following additional constraints: (a) each of the four 40-item pseudoexams contained no more than 5 items from a given passage group, (b) each of the 40-item pseudoexams contained at least 1 item from every passage group, (c) all cognates belonging to a given passage group were separated by at least 9 items from other passage groups, (d) no complete passage group appeared in the first two pseudoexams, and (e) every passage group was dispersed across at least three pseudoexams.

It is not feasible to generate multiple item orders, even with 160 items, because of the severe constraints on item proximity just described. Nonetheless, it is just these constraints, imposed to ensure wide dispersion of cognates, that make any confounding order effects unlikely.

It was not necessary to impose special constraints on "main-idea" items. These test for knowledge of the gist of a passage and hence may be especially susceptible to information from cognates. But such items are rare on the SAT–I because there is a brief overview before each passage. The five main-idea items in this study were widely dispersed across the pseudoexams. The pseudoexams were presented in two sessions, 2 days apart, as in the control condition. Two pseudoexams were given during the first session, and two during the second. For each session, 30 min was allowed per pseudoexam, with a 5-min rest between exams.

Results

Examinee Performance

Findings are based on 160 test items, and an alpha level of .05 is used throughout for declaring statistical significance. Because each item had five choices, the probability of a correct answer was 20%. Therefore, performance at the chance level would be 32 correct items if answers were chosen randomly. Examinees in both the control and quasi-random conditions performed well above that level. Mean total test score was 54.9 (SD = 12.1), or 34%, correct in the control condition, which is statistically greater than chance in a one-sample test against 20% correct items, $t(82) = 17.3, p < .001$, one-tailed; r^2 = 79%. For the quasi-random condition, mean total test score was 46.9 (SD = 10.8), or 29%, correct, also greater than chance in a one-sample test against 20%, $t(106) = 14.2, p < .001$, one-tailed; r^2 = 66%. These findings replicate those of Katz et al. (1991). However, unlike Katz et al. (1991), where no difference in performance was found between the two conditions, mean total test score was greater in the control than in the quasi-random condition (54.9 vs. 46.9), $t(188) = 4.8, p < .001$, one-tailed; r^2 = 11%. This suggests that for the SAT–I reading task, unlike the older version, cognates influence performance to some degree and that by diminishing (by quasi-randomizing) the information they provide, performance declines.

Correlation Between Examinee Performance and Verbal SAT Score

A Pearson product–moment correlation between each examinee's total score and verbal SAT score was carried out separately for examinees in this study with recentered (n = 130) and nonrecentered (n = 43) verbal SAT scores, respectively. The results revealed a strong relationship between verbal SAT and total score in both cases. For recentered scores, $r(55) = .70, p < .001$, in the control condition (r^2 = 49%), and $r(73) = .67, p < .001$, in the quasi-random condition (r^2 = 45%). These correlations do not differ significantly, $z(122) < 1.0, p > .05$. For the smaller sample of nonrecentered verbal SAT scores, $r(20) = .67, p < .001$, in the control condition (r^2 = 45%), and $r(21) = .56, p < .007$, in the quasi-random condition (r^2 = 31%). Once again, these correlations do not differ significantly, $z(35) < 1.0, p > .05$. Thus, despite a more restricted range of verbal SAT scores compared with those in Katz et al. (1991), and despite superior performance in the control versus the quasi-random condition, examinee performance was strongly and equivalently related to verbal SAT score in both conditions. The finding is contrary to the prediction that the relationship would be stronger in the control condition, where cognate information is more readily available.

Effectiveness of Randomizing

We have assumed that access to cognate information was eliminated in the quasi-random condition. The order in which items were presented in that condition makes the assumption, prima facie, a reasonable one. Items appearing early in the list provide

little or no cognate information, whereas those appearing later would seem to place considerable demands on memory if cognate information already presented is to be used to advantage. There was, moreover, a large number of items (160), and these were presented over 2 days. Katz et al. (1991) demonstrated that this method of dispersing cognates is highly effective for the older version of the SAT, where substantially fewer items were used. Nonetheless, if the procedure is not completely effective in this study, the difference in performance found between the control and quasi-random conditions may be underestimated. To determine whether this is so, we examined test item performance data as follows.

The performance of each test item in the quasi-random condition was directly compared with that of its counterpart in the control condition, thus eliminating the effects of item content. Performance here is defined as "item difficulty," that is, item proportion correct (p). If cognate information were completely eliminated in the quasi-random condition, one would expect not only that control items perform better than their counterparts in the quasi-random condition but that they do so throughout the course of the test. On the other hand, to the extent that cognate information is available in the quasi-random condition, that information could be used by the examinee only as the cognates are gradually introduced. Therefore, the effects of this gradual introduction would emerge only in the later stages of testing.

Individual t tests were carried out on p for each of the test items used in this study to determine whether the items performed differently in the control and quasi-random conditions. Overall, 42 items differed in the two conditions ($p < .025$, two-tailed). Of these, 36 items performed worse in the quasi-random condition compared with their control counterparts, and only 6 performed better. But of these 36 items, 16 appeared in the first pseudoexam, 10 in the second, 6 in the third, and only 4 in the last. (Of the 6 items that were superior in the quasi-random condition, none were found in the first pseudoexam, 1 in the second, 2 in the third, and 3 in the fourth.) A chi square analysis of the 36 items revealed a statistically significant difference between the distribution of such items across the four pseudoexams (16, 10, 6, and 4 items, respectively) and the expected distribution of 9 items per pseudoexam, $\chi^2(3, N = 36) = 9.33, p < .05$. Thus it appears that cognate information does, after all, become gradually available to examinees over the course of the experiment.

Cognate information is potent on the SAT–I reading task; its influence on performance is apparent even in the most unpromising circumstances, namely when items have been quasi-randomized. Had information from cognates been completely eliminated, the differences in examinee performance between the control and quasi-random conditions likely would have been even larger than the differences revealed here.

Correlation Reconsidered

Perhaps the high correlation between examinee performance and verbal SAT score in the quasi-random condition can be explained by the fact that cognate information becomes gradually available in that condition, as we have shown. If true, one would expect the correlation computed over the second half, or 80 items, of the experiment (where cognate information exerts a statistically detectable effect), to be significantly higher than the comparable correlation computed over the first half of the experiment (where cognate information exerts little effect). However, the correlations were both high and equivalent to each other. Using the sample of recentered verbal SAT scores in this study, $r(73) = .58, p < .001$ ($r^2 = 34\%$), over the first half of testing, whereas $r(73) = .61, p < .001$ ($r^2 = 37\%$) over the second half. For the smaller sample of nonrecentered verbal SAT scores in this study, the pattern of results was similar, though the correlations were somewhat lower. Over the first half of testing, $r(21) = .53, p < .05$ ($r^2 = 28\%$), and over the second half, $r(21) = .48, p < .05$ ($r^2 = 23\%$).

Discussion

Is performance on the new version of the SAT reading task (the SAT–I) influenced by information from cognates, that is, test items belonging to the same passage? Items were presented without the passages in either a control or a quasi-random condition, the latter to eliminate the accessibility of cognate information. Cognates did indeed influence performance. Although examinees performed better than chance in both conditions, performance was best in the control. In fact, the difference between the two conditions was probably underestimated because the influence of cognates in the quasi-random condition was not completely eliminated.

Of course these findings tell us little about how, or whether, examinees use cognates when passages are available. Only a very different study, carried out with the passages, can reveal a definitive role for cognates. It may well be that in such circumstances, cognates play no role or a weaker role than demonstrated here. Examinee self-reports when the passages are available, perhaps using the method of Powers and Leung (1995), could yield preliminary clues. But more direct process measures, particularly attentional measures such as eye movement recordings, would be the most revealing. One would expect a division of attention between the passages themselves and cognates if the latter are used by examinees.

However, from a psychometric standpoint, cognate information may, after all, be of little significance. The correlation between examinee performance and verbal SAT score in both conditions of this study was high. Hence, performance without the passages was strongly related to a proxy for performance with the passages, independent of cognate information. Thus, even in circumstances where cognate information aids performance, such information seems not to account for *systematic* differences in performance when the passages are actually available. No-passage factors other than cognate information probably are the most potent predictors of performance on the SAT reading task—factors such as prior knowledge of specific subject matter, general knowledge, and testing skills linked to the multiple-choice format (Just & Carpenter, 1987; Messick, 1988; Millman, Bishop, & Ebel, 1965). Whatever the specific factors may turn out to be, they continue to raise doubts about the construct validity of the SAT reading task.

References

ATP guide for high schools and colleges. (1985). New York: College Entrance Examination Board.

8 real SATs. (1996). New York: College Board.

Feinberg, L. (1994–95, winter). A new center for the SAT. *The College Board Review, 174,* 8–13, 31–32.

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension.* Boston: Allyn & Bacon.

Katz, S., Blackburn, A. B., & Lautenschlager, G. (1991). Answering reading comprehension items without passages on the SAT when items are quasi-randomized. *Educational and Psychological Measurement, 51,* 747–754.

Katz, S., Johnson, C., & Pohl, E. (1999). Answering reading comprehension items without the passages on the SAT–I. *Psychological Reports, 85,* 1157–1163.

Katz, S., & Lautenschlager, G. (1994). Answering reading comprehension questions without passages on the SAT–I, ACT, and GRE. *Educational Assessment, 2,* 295–308.

Katz, S., Lautenschlager, G., Blackburn, A. B., & Harris, F. (1990). Answering reading comprehension items without passages on the SAT. *Psychological Science, 1,* 122–127.

Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun H. (Eds.), *Test validity* (pp. 33–45). Hillsdale, NJ: Erlbaum.

Millman, J., Bishop, C. H., & Ebel, R. (1965). An analysis of testwiseness. *Educational and Psychological Measurement, 5,* 707–727.

Powers, D. E., & Leung, S. W. (1995). Answering the new SAT reading comprehension questions without the passages. *Journal of Educational Measurement, 32,* 105–129.

Pyrczak, F. (1972). Objective evaluation of the quality of multiple-choice items designed to measure comprehension of reading passages. *Reading Research Quarterly, 8,* 62–72.

Tuinman, J. (1973–74). Determining the passage dependency of comprehension questions in five major tests. *Reading Research Quarterly, 9,* 206–223.

Received September 6, 2000
Revision received March 8, 2001
Accepted March 8, 2001 ■