DOCUMENT RESUME

ED 385 563 TM 023 988

AUTHOR Powers, Donald E.; And Others
TITLE Will They Think Less of My Handwritten Essay If Others Word Process Theirs? Effects on Essay Scores of Intermingling Handwritten and Word-Processed Essays. The Praxis Series: Professional Assessments for Beginning Teachers (TM).
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-92-45
PUB DATE Jan 92
NOTE 30p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Beginning Teachers; *College Students; *Essays; Essay Tests; Evaluators; *Handwriting; Higher Education; Pilot Projects; *Scoring; Teacher Evaluation; Test Items; *Word Processing
IDENTIFIERS Praxis Series; *Test Rescoring

ABSTRACT

A study was undertaken to determine the effects on essay scores of intermingling handwritten and word-processed versions of student essays. A sample of 32 examinees, each of whom had produced both a handwritten and a word-processed essay, was drawn from a larger group who had participated in a pilot study of item types for The Praxis Series: Professional Assessments for Beginning Teachers (TM). Students' original handwritten essays were converted to word-processed versions, and their original word-processed essays were converted to hand-written versions. In a preliminary study, essays were intermingled and rescored. Analyses revealed higher average scores for essays scored in the handwritten mode than for those scored as word-processed, regardless of how they were originally produced. Several hypotheses were advanced to explain the discrepancies between scores on handwritten and word-processed essays. The training of essay readers was modified based on possible explanations for the discrepancy, and the experiment was repeated. The second study showed about a 25% reduction in the discrepancy. The effects in the second study were small and were predicted to have little if any impact on certification decisions. It is recommended that the scoring be monitored and that the modified training be adopted. Two tables and two figures present study findings, and the appendix contains writing samples. (Contains eight references.) (SLD)

THE PRAXIS SERIES
Professional Assessments for Beginning Teachers™

Will They Think Less of My Handwritten Essay if Others Word Process Theirs? Effects on Essay Scores of Intermingling Handwritten and Word-Processed Essays

Donald E. Powers
Mary E. Fowles
Marisa Farnum
Paul Ramsey

January 1992

RR. 92-45

Educational Testing Service

Will They Think Less of My Handwritten Essay if Others Word Process Theirs?

Effects on Essay Scores of Intermingling Handwritten and Word-Processed Essays

Donald E. Powers

Mary E. Fowles

Marisa Farnum

Paul Ramsey

January 1992

Acknowledgments

Thanks are extended to Eleanore DeYoung, Kathy Fairall, Meg Guggenbickler, and Ruth Yoder for their excellent work on coding and transcribing essays, to Kate Pashley who processed and analyzed the data, to Charlie Lewis for advice on statistical analyses, to Hunter Breland, Roberta Camp, and Kalle Gerritz for helpful reviews and suggestions on an earlier draft.

Abstract

A study was undertaken to determine the effects on essay scores of intermingling handwritten and word-processed versions of student essays. A sample of examinees, each of whom had produced both a handwritten and a word-processed essay, was drawn from a larger sample of students who had participated in a pilot study of item types being considered for the new academic skills assessments of The Praxis Series: Professional Assessments for Beginning Teachers™.

Students' original handwritten essays were converted to word-processed versions, and their original word-processed essays were converted to handwritten versions. In a preliminary study, handwritten and word-processed essays were then intermingled and rescored.

Analyses revealed higher average scores for essays scored in the handwritten mode than for essays scored as word-processed, regardless of the mode in which essays were originally produced. Several hypotheses were advanced to explain the discrepancies between scores on handwritten and word-processed essays. The training of essay readers was subsequently modified on the basis of these hypotheses, and the experiment was repeated using the modified training with a new set of readers.

The results of this second study showed an average reduction of about 25% in the discrepancy between essays read in the handwritten and word-processed modes, compared with the results of the initial study. The effects computed in the second experiment were small by most standards, and predicted to have very little if any impact on certification decisions. Nonetheless, a recommendation is made not only to adopt the modified training but also to monitor this effect throughout the operational scoring sessions.

Will They Think Less of My Handwritten Essay if Others Word Process Theirs?

Effects on Essay Scores of Intermingling Handwritten and Word-Processed Essays

In the not so distant future, familiarity and comfort with computers will undoubtedly be a less important issue in assessment than it is today. In all likelihood, most if not all students will have used computers routinely--to process reports and to manipulate data, for example. In the short term, however, not everyone who is asked to take a computer-administered test will be equally comfortable with this technology. For example, it is likely that initially not all of those who take the academic skills portion of The Praxis Series: Professional Assessments for Beginning Teachers™ will prefer to produce an essay on a computer. Therefore, all examinees will be permitted either to handwrite an essay or to use a basic word processing program to generate it. As a consequence, some essays will look very different from others. A concern is that this different appearance may translate to different scores: either handwritten or computer-produced essays may suffer when evaluated with essays produced in the other mode.

There appears to be little if any evidence that speaks directly to this concern, which may be motivated mainly by a number of studies that have consistently found a relationship between quality of handwriting and scores assigned to essays (See, for example, Markham, 1976, and reviews by Hughes, Keeling, and Tuck, 1983, and by Chase, 1986.). Some direct evidence has been presented by Arnold, Legas, Obler, Pacheco, Russell, and Umbdenstock (1990), who found that, contrary to initial expectations, student papers that were converted to word-processed versions received lower scores than did the original handwritten versions. The results of this study are discussed more fully later in this report.

The purpose of the study described here was to determine for the essays that examinees will write as part of the computer-based Praxis I: Academic Skills Assessments any effects of the mode (handwritten or computer produced) in which essays are scored.

Method

A representative sample of 32 writers was drawn from a larger sample of 568 students who had participated in the summer of 1990 in a pilot study of item types being considered for the new academic skills assessments of The Praxis Series. The characteristics of the total sample have been described elsewhere (O'Neill & Kubiak, 1991). For purposes of the study described here it should be mentioned that fewer than 5% of the sample said that they had never used a computer, and only 10% of computer users rated themselves as poor typists (90% rated themselves as either average or good).

Nearly 500 of these students produced at least two essays--one in handwritten form and one on computer. For the computer-produced version, basic word processing functions and a tutorial were provided. For each essay, students were given a pair of topics and asked to select one on which to write in the allotted 50 minutes. A total of ten topics was spiralled through the computer and paper-and-pencil testing rooms. Essay topics were of two different types. Three pairs were classified as personal experience (e.g., "Describe the most interesting course you have taken...") and the remaining seven pairs as general issues (e.g., "Should students wear uniforms in school? Discuss your position on dress codes..."). All essays were scored independently by two trained readers using holistic scoring methods to generate scores on a 1 to 6 scale. A third reader was employed to resolve discrepancies of two points or more.

The sample of 32 was drawn to represent the entire range of performance as expressed on the six-point (1 to 6) essay score scale. Selection was based on the total score, ranging from 2 to 12, previously assigned to each essay by the two readers. So as not to inadvertently introduce any bias in sample selection, half of the sample was drawn on the basis of scores on the computer-produced essay and the other half on the basis of scores on the handwritten essays. For example, the following distribution of scores was obtained from handwritten essays:

one "12", two "10"s, four "8"s, four "6"s, three "4"s, and two "2"s. The corresponding word-processed essays were then retrieved for the same students, regardless of the scores that had been assigned to the word-processed essays. For a second sample of students, the same number (16) and distribution of essays was then drawn on the basis of scores on the word-processed essays, and the corresponding handwritten essays for this sample were then retrieved without regard to scores. Essays that had received total scores of 6 or 8 were oversampled slightly, as these score levels were anticipated to be indicative of barely adequate performance. In this manner, both a word-processed and a handwritten essay were retrieved for each of 32 examinees. No attempt was made to sample topics or topic types.

The handwritten essays were then word processed so that they resembled those that were originally produced on the computer. Processors were instructed to reproduce the essays exactly as originally written and not to correct misspellings, grammatical errors, etc. In addition, each examinee's original computer-keyed essay was handwritten by one of several transcribers. Transcribers were told to correct any mistakes that were obvious typographical errors, e.g., "hte," but not to make any other changes. In order to introduce some variation in handwriting style, several transcribers were used. In addition, in order to reflect the general quality of student handwriting each transcriber first "studied" the student's handwritten essay and then attempted to emulate both the style and quality of penmanship and the general overall appearance of the essay in terms of neatness (number of erasures, crossouts, etc,).

It was clear from the start that at least one such transcription, e.g., from handwritten to computer-produced, was necessary in order to disentangle the effects of the mode in which essays were produced and the mode in which they were scored. That is, because examinees might actually create better (or worse) essays with a word processor, we would not have been able to determine whether any score differences resulted from the manner of production or

from the mode of scoring.[1] Initially, the double transcription was thought to serve as a replication primarily. However, as the study progressed it became clear that converting essays from handwritten to word-processed modes, and vice versa, was not as straightforward as we had envisioned and that subtle decisions were required to preserve the intentions of essay writers. Furthermore, because the decisions sometimes depended on the direction of the translation, the two transcriptions, i.e., from word processed to handwritten and vice versa, did not mirror one another precisely. The possibility existed that in converting the essays we might also inadvertently introduce more subtle changes, thus producing higher or lower quality essays in one mode than in the other. That is, we may have lost (or gained) something in the translation. It seemed less likely that any such bias would be the same for both types of transcription.

Each of the resulting 128 essays (32 writers by 2 modes of production by 2 conversions) was rescored by two trained readers who had not been involved in the initial scoring of the essays. These readers used the same scoring guide and standards as the original readers. The scoring was accomplished in the following manner. Two separate stacks of 64 essays were assembled so that every other essay was either handwritten or word processed. One stack contained both the original handwritten and original word-processed essays of the sample of writers; the other contained the conversions of these essays in the same order as found in the first stack. Because the content of the essays was the same in each stack, a different pair of readers was assigned to each set. Rescoring was performed in the same manner as the original scoring.

In order to assess the degree to which the transcribers were successful in emulating the style of the writers, two project staff members coded the appearance of each handwritten essay--

[1]As K. Gerritz has noted, transcribing handwritten essays does not produce computer-written essays (or vice versa), because writers may use different strategies, make different kinds of errors, etc. when producing essays in the two different modes.

both the original and the transcription.[2] A scale ranging from 1 (very untidy/very difficult to read) to 5 (very tidy/very easy to read) was used. Although this scale combined elements of both appearance and legibility, it was suggested to coders that an appropriate criterion on which to base their judgments would be how quickly they were able to read each essay. Hence, legibility probably weighed heavily in the ratings.

Results

Study 1

Rescoring of the essays revealed a high level of interreader agreement, as was noted in the earlier scoring of the much larger sample of essays. The correlation between the two readers (A and B) who scored the original essays was .87 for the handwritten papers and .93 for the computer-produced papers. The corresponding correlations between readers (C and D) who scored the converted essays were .97 and .98. Intercoder agreement on the tidiness/legibility of handwritten essays was modest.[3]

Table 1 displays the means of the actual essay scores (on the 6-point scale) assigned by each reader to essays in each of the four conditions. For example, the mean score assigned by reader A to 32 original handwritten essays (scored as handwritten) was 3.66. The mean score assigned by reader C to these same essays was 2.94 after they had been converted to word-processed versions. Figure 1 shows these results for each condition, averaged over each pair of

---

[2]Although experienced essay readers do not identify this feature as influencing the scores they award, some critics do perceive it as having a strong impact. For this reason we have chosen to consider this characteristic in this study.

[3]The correlation between coders was .61 for the original handwritten essays and .47 for the original computer-produced essays when converted to the handwritten mode. The correlation between coders' ratings of tidiness/legibility of the original handwritten essays and the essays that were converted to handwritten form by our transcribers was .48, suggesting that transcribers were moderately successful in imitating the quality of students' handwriting. (Samples of transcribers' skills are shown in Appendix A.) However, transcriptions were rated as significantly more tidy/legible on the 5-point scale than was the handwriting of students in the sample (3.42 vs 2.73), $t(31) = 3.78$, $p < .001$.

readers. As can be seen, when original handwritten essays were word processed and rescored, the average score (based on two readers) decreased significantly from 3.63 to 2.98, $t(31)=5.89$, $p<.001$. Individually, 21 of 32 scores decreased, 10 remained the same, and only one increased. When the original computer-produced essays were handwritten and rescored, the mean score increased slightly--from 3.27 to 3.45, $t(31)=1.96$, $p=.10$. A total of 11 scores increased, 16 remained the same, and 5 decreased. The average change was significantly larger (.65 vs. .18) when handwritten essays were converted to word-processed versions than when word-processed essays were converted to handwritten versions, $t(31)=3.13$, $p<.01$.

For both conversions the effect appeared larger for essays that received low scores (3 or less) than for those that received high scores (more than 3). Upon conversion to word-processed versions, low-scored handwritten essays decreased on average by .75 on the 1 to 6 scale vs. .55 for higher rated essays. Upon conversion to handwritten versions, word-processed essays increased by .22 for low-scored essays and by .14 for higher rated essays.

In order to determine possible sources of score discrepancies, writing researchers analyzed those essays whose scores changed the most when they were converted. This examination revealed, first of all, that the word-processed versions appeared to be considerably shorter than the corresponding handwritten versions. The single spacing of the printouts seemed to exacerbate this feature.

A second observation (and one that was reinforced by the comments of one of our coders, a former school teacher) was that poor handwriting at opportune times can mask a multitude of sins. For word-processed essays, on the other hand, a variety of errors (e.g., misspellings, grammatical mistakes, and inappropriate paragraphing) tend to be more readily apparent and even glaring.

Finally, it was evident from the "strikeouts" in several of the handwritten essays that some examinees had made serious efforts to revise their essays, either as they wrote or after they completed their drafts. Any such effort was, however, not visible in the word-processed versions of the essays. The possibility is open, therefore, that readers may have rewarded the effort that was implied by the revisions apparent in the handwritten essays. Because teachers of writing are often trained to encourage students' attempts at revision, this possibility was plausible.

Study 2

A priori, there was good reason to believe that modified training might affect the ratings of essay readers (see Freedman, 1981, for example). The first study was repeated therefore in the same manner with the same essays, but with four new readers, who received modified training. In brief, the major modifications entailed

(1) emphasizing that handwritten and word-processed essays may make different impressions. The results of Study 1 were discussed, and readers were encouraged to try to "get beyond" the mode in which an essay was presented. Staff made it clear, during the training, that we were trying to eliminate this effect.

(2) discussing the influence of (perceived) length on essay scoring

(3) using both handwritten and word-processed essays in the training, and

(4) checking for differences in the standards applied to scoring essays in the two modes.

Also, all word-processed essays were double spaced for this study, so that these versions would not appear to be dramatically shorter than handwritten essays.

In addition, a counterbalanced design was employed in the scoring sessions in order to control for any possible differences in standards employed by pairs of readers.[4] That is, each reader's essay stack consisted of original essays for half of the writers and converted essays for the other half. The order was counterbalanced, with two readers getting the converted essays first and two getting the original essays first. Also, two readers first evaluated essays for writers 1-16, followed by those for writers 17-3[illegible] The other two readers evaluated the samples in reverse order. Again, as in the initial reading, the correlations between scores assigned by different readers was very high, ranging from .87 to .89 for essays in each of the four conditions (i.e., original handwritten, converted handwritten, ...).

The mean scores assigned in this second study are given in Table 2 and displayed in Figure 2. The effect was .27 to .28 points for each conversion, and the direction of the effect was the same as for Study 1. That is, handwritten essays received higher scores. Although these effects are detectable [($t(31)=2.60$, $p<.05$ and $t(31)=2.31$, $p<.05$)], they can generally be considered as "small" in the context of the behavioral sciences. The effect sizes of -.21 (when converting original handwritten essays to word-processed versions) and .19 (when converting original word-processed essays to handwritten versions) correspond precisely with Cohen's (1977) definition of a small effect as .2 (compared with a "medium" effect of .5 and a "large" effect of .8). (As in the preliminary study, upon conversion to handwritten versions, word-processed essays that received scores of 3 or less increased more than did higher rated essays--.46 compared with .13. There was, however, no difference in the decreases exhibited when low

[4]Another approach to controlling for possible differences between the two pairs of readers would have been to calibrate the reader pairs by having both pairs read an identical set of essays and then adjusting for any detectable differences in standards (Braun, 1988). This approach is especially desirable when only one reader can be used. Although this approach would have been a useful refinement here, the double reading by two readers and the reading of some essays in each experimental condition by each reader was thought to be an adequate control.

and high-rated handwritten essays were converted to word-processed versions--.27 compared with .29.)

The overall effect over both conversions was on average about 25% less in this second study than in the first, presumably because of the modified training. Compared with the initial study, the second experiment revealed greater consistency with respect to the size of the effect for the two kinds of conversions. (The effects were .65 and .18 in the first vs. .27 and .28 in the second.) This may be due to the somewhat tighter design employed in the second study. Because readers were given each of the four possible combinations of essays in the second study (i.e., originally handwritten/scored as handwritten, originally handwritten/converted to word processed, etc.), differences among readers' standards for awarding scores is less a threat to the interpretation of mean score differences among the conditions in the second study than in the first.[5]

Discussion and Implications

A preliminary study revealed that essay readers gave higher scores to handwritten essays than to word-processed essays. This result was found both when examinees' essays were originally handwritten and then converted and rescored as word-processed essays, and also when original word-processed essays were converted and rescored as handwritten essays. These results suggested that essay readers may be either more lenient in scoring handwritten essays than in evaluating word-processed essays or more harsh in evaluating word-processed essays

[5]In this second study, with any possible differences between readers' standards being more adequately controlled, it appears that there may also have been an effect of mode of production. When averaged over both modes of scoring, essays that were produced with the wordprocessor were rated higher by about .27-.28 points on average than were the essays that students handwrote. Although this was not the focus of the study, it does lay open the possibility that examinees may find the word processor to be a helpful tool in writing their essays. This question of effect of mode of production is meaningful, however, only when analyses can be conducted by level of experience/familiarity with word processing. With the sample studied here, the effect appears small and positive. With examinees having more (or less) experience with word processing the effect might have been larger (or smaller).

than handwritten ones. This finding ran counter to our initial concern that readers might award higher scores to the presumably more polished, easier-to-read word-processed essays.

These results were not unlike those found in the study by Arnold et al. (1990) at Rio Hondo College (a mid-sized community college in the suburban Los Angeles area), which was brought to our attention after our study was initiated. Upon comparing some 300 previously-scored handwritten essays with word-processed versions of the same essays, the investigators found that the word-processed versions received significantly lower scores than their original handwritten counterparts. As a follow-up, the researchers intermingled word-processed essays with handwritten ones and subjected this mixed batch to rescoring. A subsequent survey revealed that most essay readers actually preferred scoring the handwritten versions over the word-processed essays, even though the former were more difficult to read.

The authors also suggested that readers seemed to have had higher expectations for word-processed essays than for handwritten ones. It was hypothesized that readers may have been more inclined to expect fully edited, polished products when essays were word processed and that they tended to forget they were reading rough drafts written under time pressure. (This phenomenon has also been noted in informal studies by the authors.)

Finally, the investigators noted what they termed a "reader empathy assessment discrepancy (READ) effect," i.e., a greater tendency to identify with the authors of handwritten essays, and thus award higher scores to them. The researchers indicated that readers reported that, with handwritten essays, they felt closer to the writer. In the words of the researchers,

> [The readers'] empathy, in part, allowed for a closer identification with the writer's individual voice as a strong and important aspect of the essay. (p.14)

Arnold et al. (1990) also reported that, by a margin of 2 to 1, readers said that they gave students the benefit of the doubt when they encountered difficulty in reading their handwriting.

Moreover, readers' comments suggested that they often mentally transformed students' texts or filled in perceived gaps when reading handwritten essays.

Finally, the Rio Hondo researchers found that score differences, favoring handwritten essays over word-processed ones, were greater both when essays were more difficult to read and when they were relatively short. In our preliminary study, the score discrepancy between handwritten and word-processed essays was greater for lower-quality essays than for higher-quality ones. That is, poorer essays suffered more when evaluated in a word-processed format (or benefitted more when read in a handwritten format). Perhaps for these lower quality essays, there were more opportunities to invoke a benefit-of-the-doubt criterion.

The discrepancy favoring handwritten essays was greater in our initial study for essays that were originally handwritten and then converted to word-processed versions than for word-processed essays that were converted to a handwritten format by our transcribers. Again, this result was consistent with a benefit-of-the-doubt hypothesis. Our transcribers produced handwritten versions that were generally neater/more legible than the original handwritten essays of examinees. Thus, it was hypothesized that there were probably fewer instances of unreadable or difficult-to-read words among the transcribed handwritten essays than among the original handwritten essays of examinees, and therefore less opportunity to give writers the benefit of any doubt. This pattern was not detected in the second study, however, and may have resulted initially from differences among readers' scoring standards.

Several possible courses of action were considered to remedy the phenomenon that was discovered in our first study. First, all word-processed essays were double-spaced in order to reduce the perception that these essays were shorter than their handwritten counterparts. This double spacing should better emulate the spacing of the handwritten essays, and possibly contribute therefore to more comparable scores.

Among several additional alternative possibilities that were envisioned when the test becomes operational were the following:

1. Advise all examinees to handwrite instead of word process their essays, since handwritten efforts are likely to yield higher scores. For several reasons this strategy did not have much appeal, especially because it would discourage many examinees from using a helpful tool, that would typically be available to them for on-the-job writing.
2. Convert all handwritten essays to word-processed versions before they are scored. This would, of course, standardize all essays, but it would involve considerable extra expense and introduce the possibility of clerical error. In addition, we ought to feel justifiably uneasy about converting examinees' essays to a mode in which we expected scores to be lower.
3. Equate scores or make other statistical adjustments in order to render handwritten and word-processed versions comparable. This strategy was also problematical, however. At the least, a number of slippery technical problems and logical dilemmas would have needed to be addressed if equating was to be a practical solution.
4. Modify instructions/training for essay readers. Both handwritten and word-processed essays could be used in the training of readers, and extra emphasis might be placed on essays of lower quality. Essay readers could also be informed of the reader empathy effect that has been noted elsewhere. This solution appeared to be the most reasonable and plausible one. Its appeal was greatest from a validity standpoint, because it seemed to address the irrelevant source of score variation in the most direct manner, and the results of our second study suggested that this approach did have an effect.

The results of the second study using the modified instructions and training revealed a smaller effect of the mode in which essays were scored, and this effect was the same regardless of the direction of conversion. With regard to practical consequences, this effect (a raw score of .27 to .28 on a 1 to 6 score scale) may not be cause for much concern, particularly because performance on the essay will be only one component in the writing score that is to be reported, and, therefore, only a small portion of the score on which decisions will be based.

Moreover, the preceding discussion assumes that the size of the effect of mode of scoring is equivalent to the magnitude of what might be termed "bias," that is, the degree to which the effect accurately represents a deviation from an examinee's true score. In fact, however, because in all likelihood the effect results both from leniency in scoring handwritten essays and stringency in scoring word-processed essays, the "bias" may actually be as little as half the size of the effect computed here. It seems unlikely that, except for essays judged to be very close to a particular score, say, an extremely strong "3" or a very weak "4," this effect would have any practical consequence. Nonetheless, periodic monitoring of this effect during essay scoring sessions would seem to be a desirable practice. Furthermore, essays that result in large discrepancies when scored in alternative modes should be collected for further inspection in order to determine better the locus of any sources of bias. This information in turn may prove useful in making further refinements in the training of essay readers.

In conclusion, an appropriate answer to the question posed in the title of this paper appears to be an unequivocal "No, they will not think less of your handwritten essay." Examinees who are not comfortable with word processing should therefore feel free to exercise their option to handwrite essays, without fear of incurring any penalty. At the same time, however, examinees who are able to word process their essays should feel free to do so, for the

possibility exists that, with the skillful use of the word processor as a writing tool, they may produce higher quality essays.

References

Arnold, V., Legas, J., Obler, S., Pacheco, M. A., Russell, C., & Umbdenstock, L. (1990). *Do students get higher scores on their word-processed papers? A study of bias in scoring hand-written vs. word-processed papers*. (Unpublished paper). Whittier, CA: Rio Hondo College.

Braun, H. I (1988). Understanding score reliability: Experiments in calibrating essay readers. *Journal of Educational Statistics, 13,* 1-18.

Chase, C. I. (1986). Essay test scoring: Interaction of relevant variables. *Journal of Educational Measurement, 23,* 33-41.

Cohen, J. (1977). *Statistical power analysis in the behavioral sciences*. New York: Academic Press, Inc.

Freedman, S. W. (1981). Influences on evaluators of expository essays: Beyond the text. *Research on the Teaching of English, 15,* 245-255.

Hughes, D. C., Keeling, B., & Tuck, B. F. (1983). Effects of achievement expectations and handwriting quality on scoring essays. *Journal of Educational Measurement, 20,* 65-70.

Markham, L. (1976). Influences of handwriting quality on teacher evaluation of written work. *American Educational Research Journal, 13,* 277-283.

O'Neill, K., & Kubiak, A. (1991, April). *The relationship between student background and attitudes about a computerized test of basic skills*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Table 1

Mean Essay Scores by Mode of Production, Mode of Scoring, and Reader

Mode in which essay was scored	Reader	Mode in which essay was originally produced: Handwritten	Mode in which essay was originally produced: Computer
Original:			
Handwritten	A	3.66 (1.12)	-------
	B	3.59 (1.13)	-------
Computer-produced	A	-------	3.22 (1.43)
	B	-------	3.31 (1.49)
Converted to:			
Handwritten	C	-------	3.44 (1.52)
	D	-------	3.47 (1.50)
Computer-produced	C	2.94 (1.39)	-------
	D	3.03 (1.43)	-------

Note. Standard deviations are in parentheses.

Figure 1
Mean Scores on Handwritten and Word Processed Essays

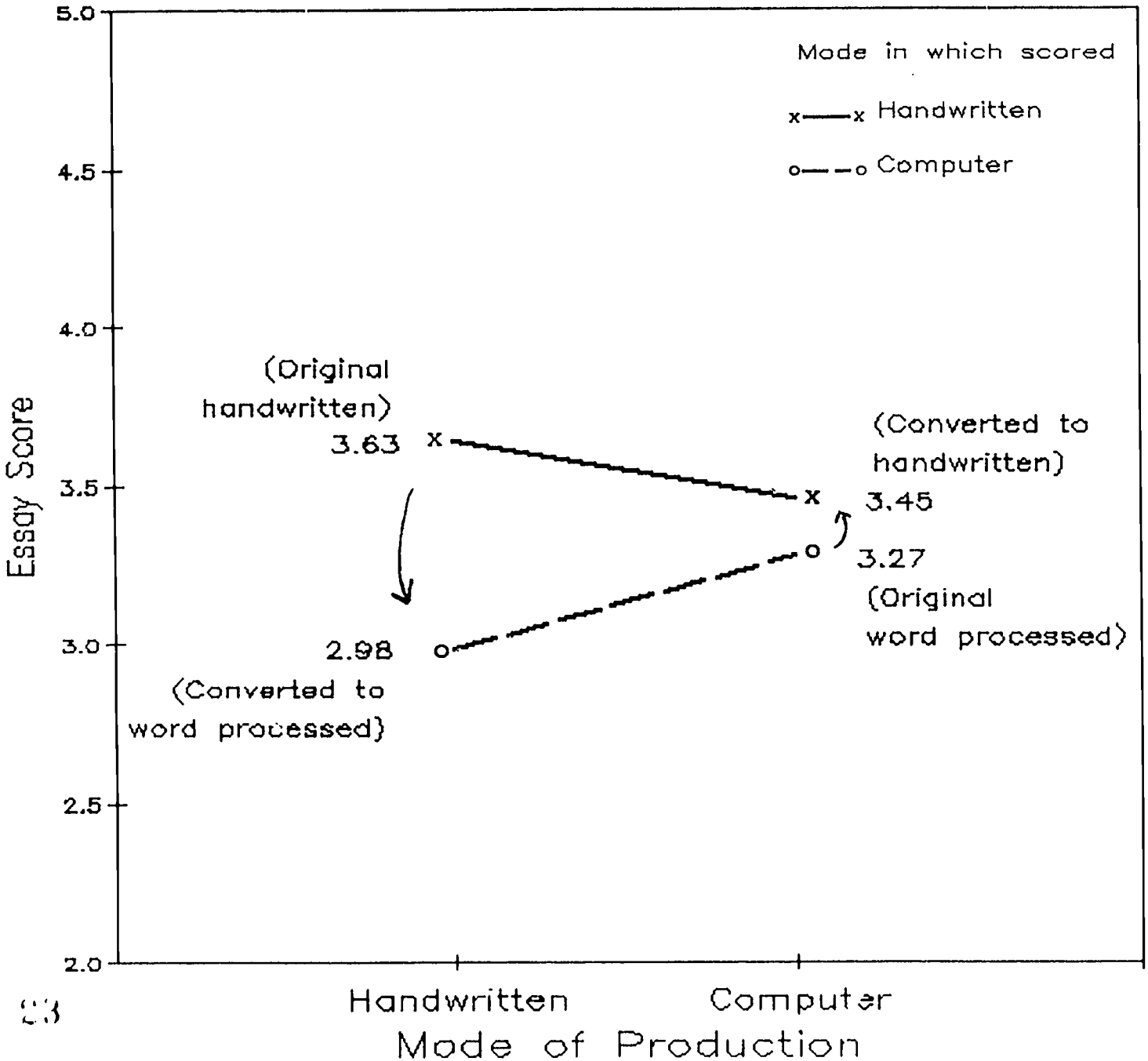


Table 2

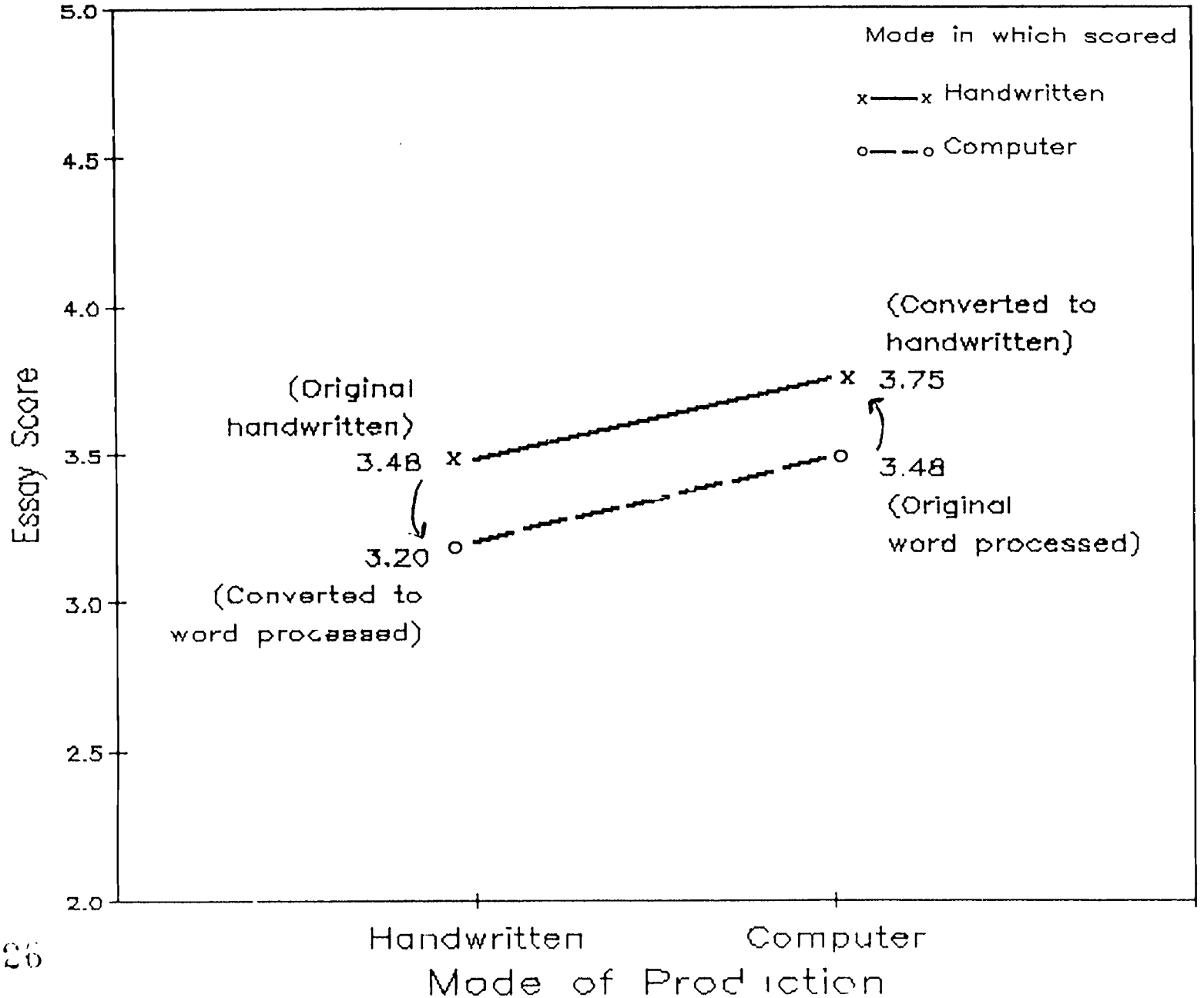
Mean Essay Scores by Mode of Production and Mode of Scoring
(Modified Training)

Mode in which essay was scored	Reader*	Mode in which essay was originally produced: Handwritten	Mode in which essay was originally produced: Computer
Original:			
Handwritten	W,X	3.56 (1.20)	-------
	Y,Z	3.41 (1.50)	-------
Computer-produced	W,X	-------	3.50 (1.35)
	Y,Z	-------	3.47 (1.50)
Converted to:			
Handwritten	W,X	-------	3.97 (1.26)
	Y,Z	-------	3.53 (1.44)
Computer-produced	W,X	3.31 (1.31)	-------
	Y,Z	3.09 (1.47)	-------

Note. Standard deviations are in parentheses.

*In this design each of the four new readers read half of the original essays (half of these being word processed and half being handwritten) and half of the converted essays (half being word processed and half handwritten). Thus the means given here are based on two readers.

Figure 2
Mean Scores on Handwritten and
Word Processed Essays
(Modified Training)



Appendix A

Comparison of Actual Handwriting and Simulated Handwriting for Three Essay Writers

Writer	Handwriting: Actual	Handwriting: Simulated
A	For instance, if a man were to enter a public area and say, "Fire!" when there really was no fire, he would create a panic and cause a clear and present danger. Similarly, if one's religion required that a person kill a person each week as a ritual, then the	Japanese students who have just grad from High School are said to be as smart an American college student in his third year accredited university, Fire. However, there are two essential facets that are being left o when this story is told. First, the fact tha
B	Though many Americans are beginning to become educated with these findings, not to many of them are trying to enforce environmental protection. There are so many different hazzards that go on every second of	In the United States need to obtain summer jobs to help then get through th school year. They also need some time to relax their brain from all the vigoro pressure that has been on it. If the students dont have time to relax that
C	Flag is a symbol of the American people as well as the American way. So, by burning the flag you are disrespecting millions of people who deserve ~~more~~ respect simple because they are people. A lot of	Moreover, I suggest that they try to keep their relationship on a lev that makes the child responsibl for what ever choice they make, let them have more freedom, and keep the communication line ope

BEST COPY AVAILABLE

53905-08580 • Y92M1.2 • 271811