Who Wrote the Book of Mormon? An Analysis of Wordprints

Author(s): Wayne A. Larsen and Alvin C. Rencher
Source: *Book of Mormon Authorship: New Light on Ancient Origins*
Editor(s): Noel B. Reynolds
Published: 1982
Page(s): 157-188

7

Wayne A. Larsen and
Alvin C. Rencher

Who Wrote the Book of Mormon? An Analysis of Wordprints

Wayne A. Larsen was Director of Advanced Research Systems, Eyring Research Institute, Inc., and is now a faculty member in statistics at Brigham Young University. He completed his undergraduate work at Brigham Young, after which he received his Ph.D. at Virginia Polytechnic Institute in 1967. His long list of publications includes articles on Minuteman II accuracy testing and advanced statistical analysis.

Alvin C. Rencher, a Professor of Statistics at Brigham Young University, also completed his Ph.D. at Virginia Polytechnic Institute. In addition to teaching, he has worked as a statistical consultant to the LDS Church, the state of Utah, and Kennecott Copper Corporation. He has published numerous articles on statistical techniques and applications in journals and magazines.

In this article, Larsen and Rencher report their findings from a statistical analysis of style in the Book of Mormon. Using "wordprint analysis," a method of determining idiosyncratic

subconscious patterns in the writings of any author, they conclude that (1) the Book of Mormon was written by many authors, and that (2) no Book of Mormon passages resemble the writing of any of the commonly suggested nineteenth-century authors. The clear yet hitherto unnoticed characteristics of the Book of Mormon discovered by Larsen and Rencher strongly support Joseph Smith's account of the book's origin.

The problem of Book of Mormon authorship has challenged historians and theologians since the book was published in 1830. Opponents of the book have claimed that Joseph Smith wrote it himself, or that an accomplice such as Solomon Spaulding or Sidney Rigdon penned it and somehow transferred it to Joseph Smith.[1] The defenders of the book maintain that it is just what it claims to be—a sacred record written on metal plates by many ancient authors and translated by Joseph Smith with divine assistance and direction (Joseph Smith—History 2:62-65).

Both sides present arguments to strengthen their case. Proponents note that proper names and cultural traits found in the book have been validated by recent Middle Eastern research,[2] while opponents point out the similarities between the book's theology and the religions of early nineteenth-century upstate New York.[3] Book of Mormon apologists find evidence of Hebrew and other ancient writing styles in the book,[4] but detractors point to the grammatical mistakes in the earlier editions as evidence that there could have been no miraculous translation.[5] Both sides also cite archaeological evidence to defend their points of view.

One element missing in all of this literature is an approach that would allow for quantification of the evidence followed by a rigorous and objective statistical analysis as a test of the competing claims. The book purports to have been written by a number of ancient authors. We can now test this claim scien-

tifically by combining certain assumptions of modern linguistics with new advances in the statistical analysis of texts.

For our analysis we started with a basic assumption that individual authors leave something analogous to a fingerprint in all their works. Each author's style has some subconscious individualistic patterns that are not easily altered. These patterns form his unique "wordprint." The growing number of wordprint studies includes inquiries into the authorship of letters, biblical books, and ancient Greek works.[6]

Stylometry

Our approach is sometimes referred to as the science of stylometry,[7] which can be defined loosely as statistical analysis of style. It is also called computational stylistics. We do not use the word *style* in the literary sense of subjective impressions characterizing an author's mode of expression. We must deal with countable items which are amenable to statistical analysis. We look then for what is frequent but largely unnoticed, the quick little choices that confront an author in nearly every sentence. Such choices become habits, so the small details flow virtually without conscious effort.

One writer on this subject, Douglas Chretien, used the term "linguistic fingerprint" to describe an author's subconscious pattern of usage of the language features which uniquely characterize his writings. He stated: "The conscious features of style can be imitated, . . . but the unconscious and subconscious features surely cannot, and a test of authorship, if it is to be reliable, must be built on them."[8]

In the literature of stylistic analysis we find many references[9] claiming that for a given author these habits are not affected by (1) passage of time, (2) change of subject matter, or (3) literary form. They are thus stable within an author's writings, but they have been found to vary from one author to another. We give two examples which illustrate this approach to authorship identification.

The first concerns the controversy over the authorship of twelve of the eighty-five Federalist Papers. Although the

Federalist Papers were first published anonymously, it was later found that five were written by John Jay and that the rest were divided between Alexander Hamilton and James Madison. Although authorship of seventy-three of the papers was determined, there was still a question as to whether Hamilton or Madison wrote the remaining twelve.

Two statisticians, Mosteller and Wallace, compared the twelve disputed papers to other of Hamilton's and Madison's writings. Using frequency of usage of the small filler words, they found overwhelming evidence favoring Madison as the author of all twelve disputed papers.[10]

As a second example, when Jane Austen died in 1817 she left an unfinished novel along with a summary. A few years ago, an anonymous admirer completed this novel and published it. She was a highly skilled author and tried her best to imitate the style of Jane Austen. She succeeded very well in the conscious elements of style but failed totally in the subconscious habits of detail. When these habit patterns were examined, the difference was clearly evident.[11]

We made the same assumption, then, that has been generally accepted and proven widely applicable: each author has a wordprint. We coined the term "wordprint" to describe a writer's linguistic fingerprint or habit patterns of usage of noncontextual words.

The noncontextual words which have been most successful in discriminating among authors are the filler words of the language such as prepositions and conjunctions, and sometimes adjectives and adverbs. Authors differ in their rates of usage of these filler words.

Some previous investigators of authorship identification have oversimplified the problem. Some have chosen a definition of wordprint and then have taken several controversial passages from an author and tested for statistically significant differences in the wordprint between passages. If any statistically significant differences occurred, they assumed different individuals had authored the passages. We believe a larger view must be taken. In addition to comparing several passages

written by the same author, we must also compare them with the works of a control group of contemporary authors. Conceivably, an individual author might produce wordprints which differ in a statistically significant manner and yet are consistent within themselves when compared with other authors' wordprints. We have taken this into consideration in our study by including authors who were contemporaries of Joseph Smith.

We propose to test the assumption that the Book of Mormon was written by one author (Joseph Smith or whomever) against the alternative hypothesis of multiple authorship. If the book were written by several people, we should statistically reject the hypothesis of single authorship. Showing multiple authorship would be strong evidence for Joseph Smith's account of the origin of the book, since it is the primary explanation which asserts multiple authors. Finding single authorship would not necessarily invalidate the believers' claims, however, because it is logically possible that even though Joseph Smith had divine direction in translating he might have paraphrased the text into his own words. This argument would also hold for Mormon's abridgment, but even then there would be other authors in Nephi and Moroni. That Joseph Smith could have received the translation word for word in a uniform literary mode with all style differences between authors obliterated is yet another possibility.

Book of Mormon Claims of Numerous Authors

According to the Book of Mormon itself, numerous prophets whose lives cover a period of over a thousand years wrote the book. Three-and-one-half centuries after the birth of Christ, Mormon realized that his writing would soon come to an end, but he was shown in vision that a later people would profit from it. Acting on divine instructions, he made a very brief abridgment of the records in his charge, engraving it on gold plates. He passed these plates on to his son Moroni, who added to the record and then deposited it in the appointed place for safekeeping. With this record compiled by Mormon and

Moroni, Joseph Smith also found a much smaller record, "the small plates," which contained the early history of these people beginning with their departure from Jerusalem soon after 600 B.C. Most of this smaller record was written by Nephi and his younger brother Jacob, who were in the original group which left the Old World. Joseph Smith used this original material in place of Mormon's abridgment covering that period. Thus, according to the text, there were four major engravers of the gold plates—Mormon, Moroni, Jacob, and Nephi—and a few minor engravers as well (see Appendix A).

In addition, the abridgers of the record often appear to be quoting from other authors; for example, Mormon recorded the commandments given by Alma to his son Helaman (Alma 36, 37). Since quotation marks do not appear anywhere in the Book of Mormon, the question remains as to whether these passages are verbatim or paraphrased.[12]

For the purpose of the statistical tests, we started with two assumptions: (1) that each of the major engravers and those they quote were distinct individuals, and (2) that the writers of each verse, or partial verse, could be identified according to information given in the text. We found very little ambiguity as to who wrote what. However, identifying the source of each verse or portion of a verse required careful scrutiny, since authorship or source shifts approximately fifteen hundred times in the text of the Book of Mormon.

Through the process of assigning each quoted segment a source, we identified over one hundred authors or originators. Twenty-two of these contributed over 1,000 words; they, along with two others who had close to 1,000 words, are listed in Appendix B in descending order according to word count.[13] As expected, Mormon is first on the list, with nearly 40 percent of the book attributed to him. Nephi has the second highest word count. The third author on this list, Alma, is not one of the engravers of the book but was quoted frequently by Mormon. A very interesting facet of this list is that if all the words attributed to Deity are combined, then Deity becomes the third most-quoted source in the book,[14] with approximately 10 percent of the words.

Non-Book of Mormon Authors

For control and comparison purposes we analyzed the writing of several nineteenth-century authors, including that of both Sidney Rigdon and Solomon Spaulding, who have been proposed as authors of the Book of Mormon. We also included other known works by Joseph Smith and contemporary works by W. W. Phelps, Oliver Cowdery, and Parley P. Pratt.[15] Also we analyzed the Lectures on Faith plus two sections from the Doctrine and Covenants. Finally we added an article called "The Paracletes," which was published anonymously in the *Times and Seasons.*[16]

Methodology

We used three basic statistical techniques: Multivariate Analysis of Variance, Cluster Analysis, and Discriminant or Classification Analysis. These techniques will be described below. We also used three basic wordprint definitions: (1) frequency of letters, (2) frequency of commonly occurring noncontextual words, (3) frequency of rarely occurring noncontextual words. Although this paper emphasizes the frequency of commonly occurring noncontextual words, all three wordprint definitions produced similar results. Appendix C contains the 38 common and 42 uncommon words we used; they were selected from a list of words ordered by frequency.

Multivariate Analysis of Variance (MANOVA)

We will first describe multivariate analysis of variance (MANOVA) and then present a few examples from the many analyses that we conducted. MANOVA is a technique that tests for homogeneity of groups,[17] the similarity of the wordprint patterns from one author to another. To illustrate the procedure, suppose that there exists a set of ten plays ascribed to Shakespeare. However, some scholars hypothesize that Shakespeare wrote only seven of the plays and that the other three were written by an unknown individual. To use MANOVA, we

divide the ten plays into two groups, one containing the seven undisputed texts, the other the three disputed plays. A wordprint definition is precisely chosen. MANOVA allows us to compare the wordprints for the two groups of plays and determines whether the observed difference in wordprint is large in relation to the internal consistency within each group of plays. A large observed difference would support the conclusion that different authors wrote the two groups of plays, while a small difference (relative to the groups' internal consistency) would suggest that one author wrote all ten plays.

Here is an oversimplified numerical example to clarify the concept further. Consider a case where we have only two authors, with three different passages from each author. We are examining the frequency of the word *and* and find the following frequency results:

	Passage 1	Passage 2	Passage 3
Author A:	.032	.031	.032
Author B:	.063	.065	.064

Frequency in this case means relative frequency; i.e., *and* appeared 32 times per 1,000 words. It is clear that, if the three selections from each author are typical, the authors will differ in the average frequency with which they used the word *and.* However, if the results were as follows, we could not discriminate between these authors on the basis of this word.

	Passage 1	Passage 2	Passage 3
Author A:	.032	.055	.068
Author B:	.042	.058	.061

On this information alone we could not rule out the possibility that A and B were the same individual.

The MANOVA technique can be applied to any number of authors and any number of words. Based on the frequencies it analyzes, MANOVA states the probability of a set of data arising if a single author wrote all of the materials examined. Certain statistical assumptions are required before this proba-

bility statement is valid. We have satisfied these sufficiently for the purposes of this study.

The writings of our 24 authors were divided into 251 blocks of text containing approximately 1,000 words apiece. Mormon was presumed to be the author of 98 of these blocks, while the last three authors—Mosiah, Enos, and the Father—had only 1 block each. The frequency of each of the words in Appendix C was computed for each of these 251 blocks.[18]

In the first analyses the blocks of words attributed to Jesus, Isaiah, and the Lord quoted by Isaiah were deleted since they agree so closely with the Bible. We thus avoid the possibility of these authors causing significant differences.

MANOVA—10 Words, Book of Mormon Only

We first compared the 21 remaining authors by using the 10 most frequently occurring words in our list. Statistically, the differences among the authors are highly significant. Differences as large as these simply could not occur if a single author wrote the book. The statistical odds that a single author wrote the book are less than 1 in 100 billion. However, this number should not be taken too literally. It depends on several assumptions, one of which is that we have a random sample of each author's writings. The 100 billion to 1 ratio does imply, however, that the authors' wordprints vary significantly with respect to each author's own internal consistency.

The 10 words which we compared were *and, the, of, that, to, unto, in, it, for,* and *be.* Only one word, *in,* was not significantly different across the 21 authors. Seven of them were significant at less than the .0001 level; i.e., the probability that a single author would produce such disparate results is less than 1 in 10,000. In a typical research study, a difference would be labeled significant if its probability level was .05 (less than 1 in 20) or smaller. Most of the differences we found were so large that the associated probability level was very much smaller than .05.

MANOVA—38 Words, Book of Mormon Only

The MANOVA was repeated using the 38 frequently occurring words listed in Appendix C, with similar results. Thus the 21 authors do not appear to be the same individual. We have not shown statistically the existence of 21 distinct styles but have strongly demonstrated wide divergence among most of the 21. The pattern of differences among the authors will be examined further in connection with the MANOVA which includes non-Book of Mormon authors as well.

MANOVA—Other Book of Mormon Tests

The preceding analyses were repeated using the Book of Mormon authors in a variety of contexts. These include analyses on word frequencies, analyses on all 24 authors (Jesus, Isaiah, and the Lord as quoted by Isaiah added to the data base), analyses on the 42 uncommon words listed in Appendix C, and analyses on frequency of letters. The results were the same in each case. We consistently found extremely low probabilities that the differences among these 24 groups of text could have been produced by a single author. There were no contradictory results.

MANOVA—38 Words, Including Non-Book of Mormon Authors

We also compared the writing in the Book of Mormon with that of Joseph Smith and his contemporaries, who wrote in the time period when the Book of Mormon was published. The 90 blocks of words we used were from Joseph Smith, W. W. Phelps, Oliver Cowdery, Parley P. Pratt, Sidney Rigdon, Solomon Spaulding, the article "Paracletes," excerpts from the Doctrine and Covenants, and the Lectures on Faith. It has been suggested that certain of these men were the authors of the Book of Mormon.

As a control test we first performed a MANOVA using all 38 words on 341 word blocks from the 33 authors (24 Book of

Mormon plus 9 non-Book of Mormon authors). Probability that differences as large as those observed could occur by chance is less than 1 in 10 billion.

The overall MANOVA results for all 33 authors is of less interest than making pertinent comparisons among the 33 authors. These comparisons include direct comparisons of the Book of Mormon and non-Book of Mormon authors, along with comparisons among the book's authors grouped appropriately. The major conclusions from these statistical comparisons are:

1. There is some evidence of a wordprint time trend within the Book of Mormon; i.e., writers are more similar to their contemporaries than to writers in other time periods. This needs further investigation.
2. The passages quoting the Father do not differ from the combined passages quoting the Lord and Jesus. But there may be a little difference between quotations from Jesus and those from the Lord.
3. There is no statistical difference between the Isaiah passages and the Lord as quoted by Isaiah.
4. Joseph Smith's writing is very different from that of the author of Lectures on Faith (see Appendix E).
5. The most salient result, however, was that *none of the Book of Mormon selections resembled the writing of any of the suggested nineteenth-century authors.*[19] The Book of Mormon itself offers the strongest evidence for a clear scientific refutation of the theories that it was written in the nineteenth century.

The MANOVA tests have shown conclusively that (1) the 21 major groups of Book of Mormon text we examined were indeed written by several distinct authors, who were individually consistent as suggested in the book itself, and (2) none of the modern candidates whom we tested for Book of Mormon authorship wrote any of that text. This leaves Joseph Smith's account as the only explanation consistent with these clear yet hitherto unnoticed characteristics of the Book of Mormon. The only alternative would be that, in spite of its growing reputation in scientific circles, the theoretical basis of wordprint is not

generally valid. But our own results on known nineteenth-century authors provide strong support for the wordprint concept.

To avoid the possibility that our MANOVA results might be unconsciously biased by any particular statistical technique, we included two additional analyses: cluster analysis and discriminant or classification analysis.

Cluster Analysis

Cluster analysis takes a series of measurements on a set of observations and identifies which observations are closest to each other. In this study, the series of measurements would be the frequencies of the 38 words which form the wordprint profile, and the set of observations would be the 1,000-word blocks. "Closeness" is defined by a distance measure of the difference between two wordprints.[20] Cluster analysis can be used as an additional test of multiple authorship, but, more importantly, it can also be used as an informal method of assessing relationships between blocks of words.

The major cluster analyses we performed yielded conclusions similar to the MANOVA results discussed earlier. Mormon's word blocks clustered with other blocks by Mormon, Nephi's with Nephi's, King Benjamin's with King Benjamin's, etc. These results were the same no matter which definition of wordprint we selected—letters, common words, or uncommon words. The percent of clusterings corresponding with the multiple authors as named in the Book of Mormon was much higher than could have been produced by chance. Since these results are very similar to those presented in the MANOVA sections, we include only two examples which show a different application of clustering.

Cluster Analysis—
24 Book of Mormon Authors

This cluster analysis was for the 24 Book of Mormon authors using one observation consisting of each author's total

words combined. Frequencies of the 38 common words were used as data. The purpose in combining each author's words was to determine how the authors relate to each other. To calculate a distance measure which would most clearly distinguish the authors, we chose the 9 words which discriminated best in the MANOVA.

Some results indicating that contemporaries write alike were—

1. Nephi's word blocks paired with those of his father, Lehi; together these then clustered with the group of word blocks of Nephi's brother Jacob and of Isaiah, the prophet most quoted by Nephi and Jacob.
2. The Lord's word blocks grouped with Jesus'.
3. Alma's word blocks grouped with those of Amulek, his missionary companion; once combined they paired with those of Abinadi, the man who converted Alma's father.
4. Samuel the Lamanite's word blocks paired with those of Nephi, son of Helaman. Samuel the Lamanite and Nephi were contemporary prophets.
5. The word blocks of the Lord as quoted by Isaiah paired with the Father's.

Some contrasting results were—

1. Mormon's word blocks paired with Helaman's, a bridge of 300 years.
2. Moroni's word blocks paired with Zenos's even though these two authors were most widely separated in time. Overall, Moroni's word blocks clustered less "correctly" than other authors'. Perhaps this is because much of his writing is an abridgment of the Jaredite record or quotation from unspecified earlier sources.

Cluster Analysis—Book of Mormon and Non-Book of Mormon Authors Combined

All 33 authors were used in this analysis, with one replication per author which consisted of all blocks combined for that

author. As before, 9 selected words were used for the distance calculations.

The following results were noted:

1. Joseph Smith's word blocks combined with those of Lectures on Faith; this pair then combined with Oliver Cowdery's (see Appendix E).
2. Jacob's word blocks combined with those of "The Paracletes."[21]
3. Nephi's word blocks combined with Lehi's.
4. Phelps's word blocks and Pratt's combined.
5. The word blocks of the Lord and Jesus combined.
6. Alma's word blocks, Amulek's, and Abinadi's combined.
7. Ammon's word blocks and General Moroni's combined.
8. Samuel's word blocks and those of Nephi (the son of Helaman) combined.
9. The word blocks of the Lord as quoted by Isaiah and those of the Father combined.
10. Mormon's word blocks and Helaman's combined.
11. Moroni's word blocks and Zeniff's combined.

In general, word blocks of Book of Mormon authors clustered with those of Book of Mormon authors, and word blocks of non-Book of Mormon authors clustered with those of non-Book of Mormon authors. The tendency of contemporaries to combine was also evident.

Discriminant or Classification Analysis

The third and most powerful statistical technique used in this study was discriminant analysis. This procedure reduced the dimensionality of differences among authors. The MANOVA has established the existence of significant differences in wordprints from one author to another. However, these wordprints are essentially 38-dimensional profiles; i.e., they are composed of the frequencies of 38 words. With 38 words to consider, it is difficult to grasp the pattern of separation between two or more authors. The discriminant procedure

determines a set of functions (fewer in number than 38) which reveal the configuration of separation among the authors.[22]

A discriminant analysis is often followed by a classification analysis in which the profile of word frequencies (wordprint) of a block of words is compared to the average profile of each author, and the block of words is assigned to the most probable author. The comparisons are made by means of classification functions which measure how closely one profile matches another. We consider the techniques of discriminant and classification analysis to be the most powerful because they are self-verifying; i.e., the results tell how well the wordprint concept works on the data being studied.

Discriminant Analysis—2000-Word Blocks for 21 Authors

The discriminant analysis we used was performed in steps. The word which best separates authors was entered first, the second best word next. This process continued sequentially until a designated critical level was reached, after which no more words were included in the analysis. In this case 18 words provided a high percentage of the discriminating power of the 38 words, and the amount of computation was thereby reduced without sacrificing much accuracy.[23] We evaluated and plotted the discriminant functions for each block of words, thus providing a visual display of the differences among authors. Some of these plots will be shown (see Figures 1 and 2).

The words selected in this discriminant analysis were then used in a classification analysis as described above. In this phase each block of words was classified with the author whose wordprint it was closest to. The percent of the correct "hits" is a measure of how well the authors can be separated, of how unique the profile of word frequencies is for each author.

In the computer run with 2,000-word blocks and 18 words selected, 93.3 percent of the blocks were correctly classified. This is a very high success rate for a situation such as this where the number of groups (authors) is so large. Typically the percent

of correct classifications drops off when the number of groups exceeds four or five, and in many applications the percentage of hits is low even when the number of groups is small. The 93.3 percentage in this case was unexpectedly high.

A better method of classifying the blocks of text is to drop one or more blocks of words from the analysis, compute the classification functions, and use these new functions to classify the blocks dropped, thus eliminating the partial circularity of the previous test. This was done on the above data base and in many other cases. The results, though not as impressive as the 93 percent just mentioned, were consistently in the 70 and 80 percent range, still very high percentages for so many groups. We performed many more analyses of this type with similar results. We mention a few.

Discriminant Analysis—
Non-Book of Mormon Authors Included

Four Book of Mormon authors who had fewer than 2,000 words were deleted. This left 162 blocks of words by 29 authors. The first two discriminant functions (see Appendix F) were evaluated for all 162 observations and are shown in Figure 1. The Book of Mormon authors are rather widely separated from the non-Book of Mormon group. It should be remembered that this two-dimensional plot is essentially a projection of higher dimensional points onto a plane. The actual points in a higher dimensional space are even more separated than they appear here.

Taken together, these tests strongly reinforce previous conclusions that—

1. distinct authorship styles can be readily distinguished within the Book of Mormon, and
2. the nineteenth-century authors do not resemble Book of Mormon authors in style.

The pattern of separation which can be noticed in Figure 1 suggests another interesting observation. The 9 non-Book of Mormon authors are known to be different. Yet their pattern of

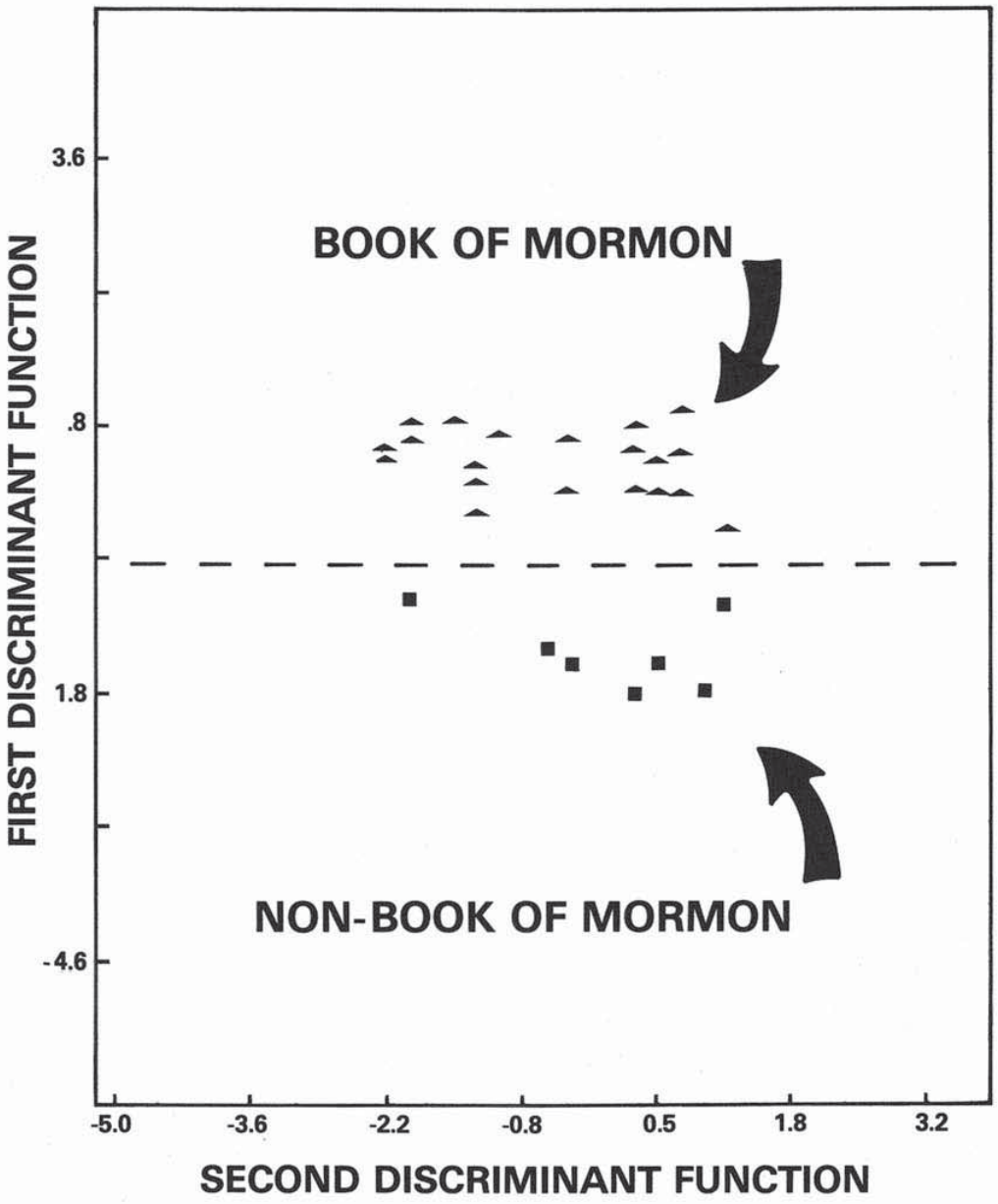

Figure 1. Discriminant Analysis of Book of Mormon and Non-Book of Mormon Authors

variation one from another is similar to the pattern of variation among the Book of Mormon authors. This emphasizes the differences among Book of Mormon authors and helps clarify that the differences we have found are neither—

1. artifacts of the book which might possibly be typical of other books, nor
2. natural random fluctuations of word frequencies from one section of the book to another.

The presence of Isaiah among the Book of Mormon authors yielded a similar result. Believers and nonbelievers agree Isaiah is a different author than the author(s) of the rest of the Book of Mormon, yet none of our statistical tests showed Isaiah to particularly stand out. That is, Mormon, Nephi, and others appeared to be as distinctively individual as Isaiah. If Joseph Smith or any other nineteenth-century author had written the book, this would not be expected.

Discriminant Analysis of Four Major Book of Mormon Authors and Joseph Smith

The intent in this analysis was to focus on the four major authors who together account for 62.2 percent of the Book of Mormon. These authors are Mormon, Nephi, Alma (the son of Alma), and Moroni (see Appendix B). These four were compared with each other and with Joseph Smith. Some 91 blocks of 2,000 words were available. Words of the King James Version were excluded, and 18 words were selected in the stepwise phase. We used four discriminant functions.

A plot of the first two discriminant functions is given in Figure 2. The following conclusions are apparent from the plot:

1. Alma's writing is different from Mormon's. Since *all* of Alma's words are taken from Mormon's writings, we can conclude that Mormon copied directly from Alma's writings and *Joseph Smith translated literally from Mormon's writings.*
2. Joseph Smith's writing is very definitely distinct from that of the authors in the Book of Mormon.

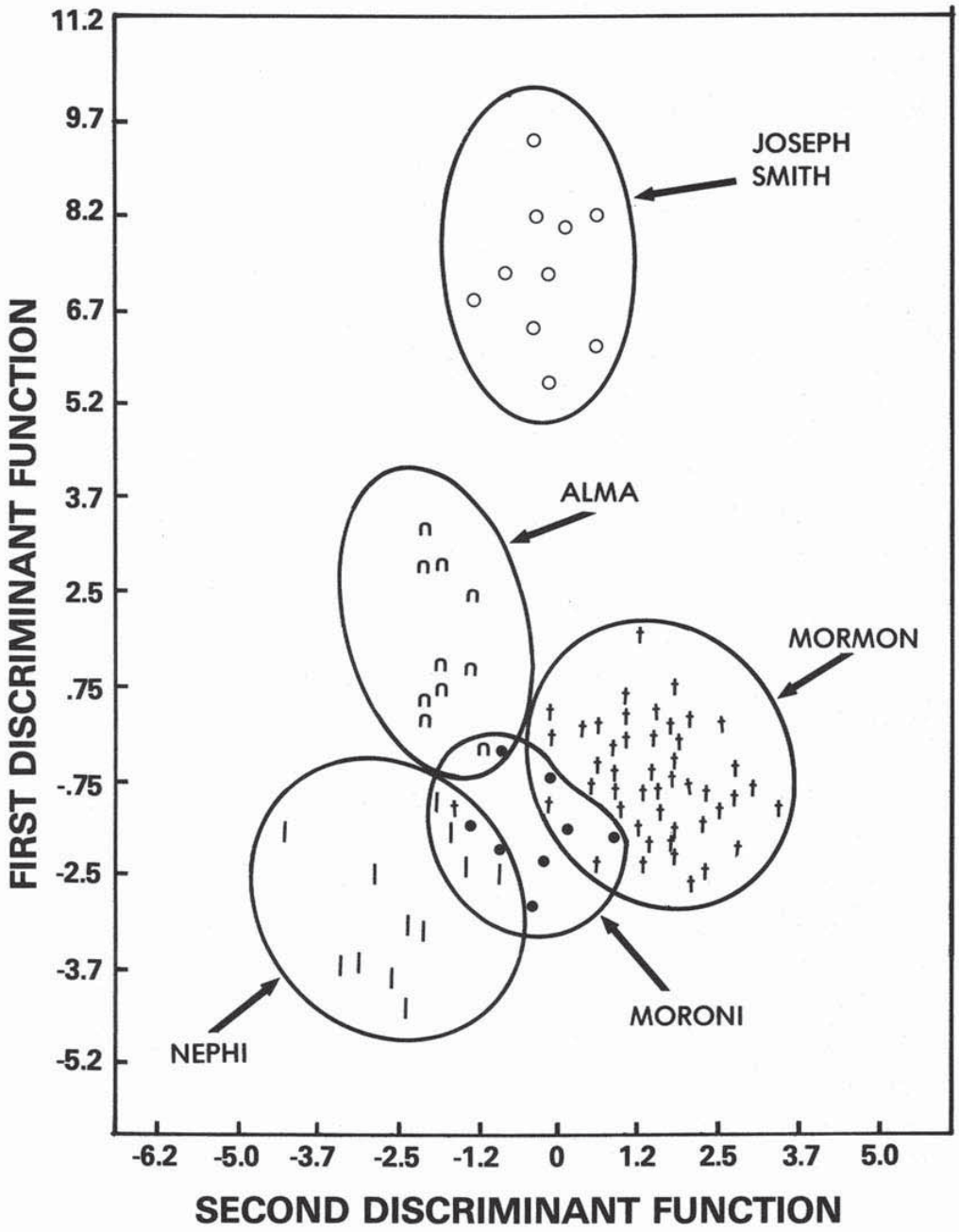

Figure 2. Discriminant Analysis of Four Major Book of Mormon Authors and Joseph Smith

3. Moroni's position between Alma, Nephi, and Mormon again indicates that Moroni is consistently hard to classify.

In the classification phase, 96.7 percent of the word blocks were correctly classified. This number speaks for itself.

Three Questions

There are three questions that may have occurred to our readers.

1. *Could Joseph Smith have altered his wordprint habits by trying to imitate the King James style?*

From all the research results with which we are familiar, the answer is no.

We mentioned the case of the lady who recently tried to imitate Jane Austen but whose own wordprint showed through the imitation when subjected to stylometric analysis. In a number of other cases, it has been shown that where an imitation is compared to the wordprint of the original, "the result resembles its creator more than it does the model."[24]

2. *Could the large differences among authors in the Book of Mormon be misleading; i.e., could we find similar differences among several works by the same author?*

In all the studies we are aware of either no significant differences were found or at most very few minor differences. As near as we can determine, the answer to this question is also no.[25]

We elaborate with a few interesting examples. One of the authors assisted in an analysis of wordprint in the Book of Isaiah.[26] Although virtually all the higher critics believe Isaiah is the product of two or more distinct authors, the Adams and Rencher work pointed to a unity of the Book of Isaiah. In fact, it showed a greater internal consistency for Isaiah than any other Old Testament book of that approximate time period.

The unity of some of Shakespeare's plays has also been questioned, but when these plays were subjected to wordprint analysis, no significant variations in wordprint were found within the given plays. An attempt to prove that part or all of Shakespeare's works were really written by Bacon resulted in

what was described by A. Q. Morton as "one of history's finest examples of serendipity."[27] A man by the name of William Friedman was hired by a prominent Baconian to unravel the ciphers or code which would reveal the identity of Bacon in the text of Shakespeare. Friedman's study actually refuted the cipher idea in Shakespeare. But he became intrigued with ciphers and went on to publish some very important papers on decipherment. His work led directly to cracking the Japanese naval code in World War II.[28]

Another study examined two books by Sir Walter Scott, one written early in his career, the other just before he died. Even though Scott had suffered four strokes during the intervening time period, there were no significant differences in wordprints either within the two works or between them.[29]

3. *Can wordprints survive translation?*

A recently completed study indicates that the answer to this question is yes. (The study was conducted by Karl S. Black, Alvin C. Rencher, and Marvin H. Folsom, with no published report yet available.) Twelve German novellas, written by twelve distinct individuals, were all translated by the same American author. When the wordprints of the twelve German authors were compared by MANOVA, differences were readily apparent, with statistical significance of a very high order.

A sizable body of writing in English by the translator was also available. When his wordprint in these writings was compared with the wordprints of the twelve German authors (translated) the differences were highly significant.

As an additional check on question 2 above, the translator's own writings were divided into subgroups. These subgroups of blocks of words were compared statistically by use of MANOVA. No significant differences were found.

Conclusions

Subject to the usual statistical assumptions and allowance for error, we make the following conclusions:

1. The wordprint hypothesis appears to be justified. Based on our analysis of known non-Book of Mormon authors, each

writer appears to have a unique set of unconscious style characteristics. This profile of usage habits can serve in many cases to identify a piece of writing as belonging to a particular author, just as a fingerprint or voiceprint can be traced to its owner or originator.

2. The results of MANOVA, discriminant analysis, and cluster analysis all strongly support multiple authorship of the Book of Mormon. According to some of the MANOVA results, the odds against the Book of Mormon having a single author are more than a billion to one. Of course the assumptions for MANOVA should be checked. For example, it is unlikely that the data can be considered to have come from a multivariate normal distribution. However, we used the arc sine transformation, which partially compensated for the lack of multivariate normality.

However, the conclusion of multiple authorship does not rest on the significance tests alone. One of the most telling arguments is provided by the plots of discriminant scores in which the variation among known authors such as Joseph Smith, Sidney Rigdon, Parley P. Pratt, and others is seen to be very similar to the variation among Book of Mormon authors. Thus if one questions the highly significant results of the MANOVA by suggesting that the differences may be statistically significant but possibly reflect only minute real differences, we can clearly refer to the graphs of discriminant functions to show that the differences among Book of Mormon authors are of the same magnitude as the differences among known authors.

Conversely, the MANOVA results reinforce the discriminant function plots. These plots exhibit a very convincing pattern of separation among authors. With the backup of significance tests, this separation becomes very real and there remains little doubt of its validity.

In further support of the MANOVA results, it should be noted that most of the 38 words were individually significant; i.e., the authors differed from each other on each word considered separately.

This finding of multiple authorship has several implications.

1. It does not seem possible that Joseph Smith or any other

writer could have fabricated a work with many discernible authorship styles (wordprints). The 24 authors do not appear in 24 separate blocks of connected words but are shuffled and intermixed in a very arbitrary manner. How could any single author keep track of 38 (actually more than 38) word frequencies so as to vary them not only randomly from one section to another but also according to a fixed underlying pattern, particularly more than a century before scholars realized that word frequencies might vary with authors?

2. The implications for translation are that the process was both direct and literal and that each individual author's style was preserved. Apparently Joseph Smith was required to render the book in a rather precise format with minimum deviations from the original "wordprint." The demonstrated presence of distinguishable authorship wordprints in the Book of Mormon argues for a formal translation in which information was transferred but the imprint of the original language remained.

3. The Book of Mormon authors taken individually or collectively do not resemble any of the nineteenth-century authors which we considered, taken individually or collectively. These authors include Joseph Smith and his contemporaries who have been considered as possible contenders for authorship of the Book of Mormon. The overwhelming evidence given by MANOVA and discriminant analysis, and to a lesser extent by cluster analysis, should discredit the alternative theories that Joseph Smith, Solomon Spaulding, or others wrote it.

The separation between Book of Mormon and non-Book of Mormon authors was established by both MANOVA and discriminant analysis. Especially convincing were the plots of the first two discriminant functions. In these plots the two groups could be cleanly separated by a straight line, an extremely rare occurrence in discriminant analysis studies. This visual separation was confirmed by the MANOVA significance test, and the possibility that the observed pattern was a chance arrangement was thus ruled out.

4. An analysis of letter counts (not detailed in this paper) yielded similar results to the word count data. Letters are obviously a rough way of detecting a wordprint, since many

contextual words contribute to the letter count. The method, however, seems to be fairly effective.

5. In a cluster analysis including both Book of Mormon and non-Book of Mormon authors, the Book of Mormon authors clustered with themselves, and the nineteenth-century authors clustered with themselves.

6. Each of the discriminant analyses was followed by a classification analysis, wherein each block of words was classified according to which author's wordprint it most resembled. When all the blocks of words were used in computing the classification functions and then submitted one by one for classification, the percentage of correct classifications varied from 69 to 100. When one block at a time was withheld from computation and then submitted, the percentage of correct classifications varied from 50 to 81 percent. These percentages are rather high considering the number of authors being classified and, therefore, reinforce the multiplicity of authors conclusion shown by the MANOVA and discriminant analysis.

7. An analysis was done using 42 words which were not among the 38 words used in the previous analyses. These 42 words occurred less frequently than the 38. The MANOVA results also showed the Book of Mormon authors differ from each other in their rates of usage of these words. In fact, the indicated level of significance showed the differences to be even more highly significant than those determined with the 38 words.

The evidence to date is that many authors wrote the Book of Mormon.

APPENDIX A
Number of Words by Engravers

Engravers	Words	Percent of Book
Mormon	174,610	65.1%
Nephi	54,688	20.4%
Moroni	26,270	9.8%
Jacob	9,103	3.4%
Enos	1,157	.4%
Amaleki	919	.3%
Jarom	731	.3%
Omni	160	.1%
Amaron	154	.1%
Abinadom	96	.0%
Chemish	69	.0%

APPENDIX B
Major Book of Mormon Writers

Author	Words	Percent of Book
Mormon	97,777	36.5%
Nephi	29,320	10.9%
Alma II	19,777	7.4%
Moroni	19,408	7.2%
Lord	12,200	4.6%
Jesus	9,654	3.6%
Jacob	8,493	3.2%
Isaiah	6,478	2.4%
Helaman	5,121	1.9%
Lehi	4,634	1.7%
Lord (quoted by Isaiah)	4,355	1.6%
Zenos	4,230	1.6%
Benjamin	4,204	1.6%
Amulek	3,158	1.2%
Samuel the Lamanite	3,068	1.1%
General Moroni	2,970	1.1%
Abinadi	2,767	1.0%
Ammon	2,417	.9%
Nephi (Son of Helaman)	2,214	.8%
Angel 1	2,083	.8%
Zeniff	1,811	.7%
Mosiah	1,167	.4%
Enos	967	.4%
Father	961	.4%

APPENDIX C
Frequently Occurring Noncontextual Words

Word	Number of Occurrences	Word	Number of Occurrences
the	20015	with	1520
and	16669	yea	1245
of	11838	should	1180
that	6883	by	1201
to	6488	as	1048
unto	3642	upon	1080
in	3705	but	991
it	3100	also	1048
for	2524	from	1007
be	2513	there	820
which	2238	because	799
a	2233	these	749
not	2090	therefore	663
came	1644	when	632
pass	1525	if	648
behold	1634	even	689
all	1788	into	686
this	1454	would	612
now	1230	forth	609

Infrequently Occurring Noncontextual Words

Word	Number of Occurrences	Word	Number of Occurrences
out	591	about	262
after	507	must	244
among	582	then	224
against	557	every	227
thus	478	what	179
according	528	nevertheless	178
again	479	until	202
may	515	exceeding	175
no	474	thereof	149
wherefore	419	through	115
before	436	towards	101
might	464	verily	76
or	438	notwithstanding	67
on	420	whatsoever	72
at	397	lest	75
away	381	whether	49
an	389	nay	44
so	358	ever	36
over	323	whereby	26
O	264	thereby	37
could	281	between	32

We thank Charles Bush for these word counts which correct those published in the earlier version of this paper.

APPENDIX D
Miscellaneous Tests Internal to the Book of Mormon

We comment briefly on two questions we tried to resolve using MANOVA. The first question involves the unity of Isaiah. Many present-day Bible scholars accept the theory that there were at least two authors of the Book of Isaiah. The principle divisions are chapters 1-39 and 40-66. We compared these two using word frequencies for the portions available in the Book of Mormon. Although we ran this test four times, we could get no significant results. This means we were unable to detect any statistical difference which would support the theory that Isaiah has more than one author.

The Sermon on the Mount as recorded in Matthew was compared with Jesus' teachings to the Nephites as recorded in 3 Nephi *excluding* chapters 12-14 which contained material similar to the Sermon on the Mount. There were 2 replications (1000-word blocks) for the Sermon on the Mount in Matthew and 7 for Jesus in 3 Nephi. Due to the small number of blocks it was necessary to run 5 analyses of 4 words each. Only 1 of the 5 tests achieved a probability level as low as .05. Thus there is little evidence of a style disparity between Jesus in the New Testament Sermon on the Mount and Jesus in 3 Nephi (excluding Sermon on the Mount material).

Again, a word of caution is needed. The tests on Isaiah and Jesus involved much smaller sample sizes than the tests on the book as a whole; therefore statistical differences would be harder to find, even if there were a real difference.

APPENDIX E
Lectures on Faith

Who Wrote the Lectures on Faith?

Most Latter-day Saints attribute the Lectures on Faith to Joseph Smith. However, historians have long been doubtful of this identification, since the lectures were originally published unsigned. Recently Alan J. Phipps completed an authorship study on the Lectures on Faith.[30] Our conclusions largely support his results, with some differences as described below.

First a cluster analysis was performed on the 9 non-Book of Mormon authors. The Lectures on Faith paired with the writings of Sidney Rigdon—which is the same general conclusion that Phipps made.

Discriminant Analysis, Non-Book of Mormon Only

In this analysis each of the 7 lectures of the Lectures on Faith was counted as 1 block (there were 7 blocks for 7 lectures).

The computation set consisted of 7 non-Book of Mormon authors with 36 blocks of 2000 words. Eight words were used as dependent variables and 4 discriminant functions were retained.

A plot of the first two discriminant functions shows 6 out of the 7 lectures grouping with Sidney Rigdon's known writings. There is no overlap of this group with other writers. The fifth lecture is rather distant from this group and is somewhat closer to W. W. Phelps's group. The fifth lecture has only 772 words, which may not be sufficient for a stable estimate of word frequencies.

In the classification phase, 88.9 percent of the blocks from the computation set were correctly classified. The lectures of the Lectures on Faith were classified as follows.

Lecture	1st Choice Author	Probability	2nd Choice Author	Probability
1	S. Rigdon	1.0		
2	J. Smith	.524	S. Rigdon	.339
3	S. Rigdon	1.0		
4	S. Rigdon	.988	J. Smith	.005
5	W. W. Phelps	.461	P. P. Pratt	.367
6	S. Rigdon	1.0		
7	S. Rigdon	.995	J. Smith	.005

These results differ somewhat from Phipps's conclusions. He assigned Lectures one and seven to Sidney Rigdon and five to Joseph Smith. He claimed that Lectures two, three, four, and six possessed elements of both men's style and concluded that these four represented a collaborative effort.

APPENDIX F
Standardized Discriminant-Function Coefficients

Word	Function 1	Function 2
and	−0.35	0.15
the	0.04	0.42
of	−0.21	−0.14
that	−0.11	−0.24
to	−0.09	0.25
unto	−0.21	−0.10
in	0.07	−0.14
it	−0.01	0.16
for	−0.51	0.15
be	0.08	−0.28
which	−0.08	−0.01
a	0.05	0.11
this	0.01	−0.29
now	−0.05	0.07

with	−0.02	0.19
upon	0.04	−0.10
but	0.05	−0.02
from	0.05	0.04
therefore	−0.11	−0.24
even	−0.07	0.03

These are the coefficients for a weighted average. Thus Function 1 = $-.35Z_1 + .04Z_2 - .21Z_3 - \ldots - .07Z_{20}$ where the Z's are the standardized frequencies of the words. The sizes of the coefficients are related to their importance in separating the authors. In Function 1, the words *and, of, unto, for,* contribute heavily. In Function 2, the most important contributors are *the, that, to, be, this,* and *therefore.*

APPENDIX G
Further Questions

The study reported here is the first major computer analysis of its kind that we are aware of. It raises a number of questions for further study which we list here.

First, we need to devise better definitions of wordprints using, for example, phrases as well as words. "And it came to pass that" was undoubtedly one word in Reformed Egyptian. Conversely, some words with two or more distinct meanings should be separated in wordprint definitions.

Second, we need to determine whether the discriminant functions possess any intrinsic meaning. An investigation of this in conjunction with more precise definitions of wordprint might be particularly fruitful.

Third, we need more investigation of wordprint time trends. In particular, the Jaredite record should be compared with the rest of the book.

Fourth, we need to take a closer look at why Moroni was relatively poorly classified.

Fifth, we need to determine what differences are introduced by using the 1830 edition of the Book of Mormon rather than the present edition.

Finally, we need to determine whether some of the misclassifications are correct after all. For example, from the context of Alma 29 it is clear that Alma is writing, yet Mormon does not identify this as a quotation. This is the only instance we found of this nature. Did we miss some others? A careful misclassification study might yield some light on this subject.

NOTES

1. See Lester E. Bush, Jr., "The Spaulding Theory Then and Now," *Dialogue: A Journal of Mormon Thought* 10(1977):40-69, for an excellent summary.

2. See Hugh Nibley, *An Approach to the Book of Mormon*, 2nd ed. (Salt Lake City: Deseret Book Co., 1964).

3. See Thomas F. O'Dea, *The Mormons* (Chicago: Univ. of Chicago Press, 1957), and Eber D. Howe, *Mormonism Unvailed [sic] or a Faithful Account of That Singular Imposition and Delusion from Its Rise to the Present Time* (Plainsville, Ohio: n.p., 1834).

4. See John A. Tvedtnes, "Hebraisms in the Book of Mormon: A Preliminary Survey," *Brigham Young University Studies* 11(Autumn 1970):50-60, and John W. Welch, "Chiasmus in the Book of Mormon," *BYU Studies* 10(Autumn 1969):69-84.

5. See Brodie Crouch, *The Myth of Mormon Inspiration* (Shreveport, La.: Lambert Book House, 1968).

6. Some of these studies are Glade L. Burgon, "An Analysis of Style Variations in the Book of Mormon," M.A. thesis, Brigham Young University, 1950; Alan J. Phipps, "The Lectures on Faith: An Authorship Study," M.A. thesis, BYU, 1977; L. LaMar Adams and Alvin C. Rencher, "A Computer Analysis of the Isaiah Authorship Problem," *BYU Studies* 15(Autumn 1974):95-102; L. LaMar Adams and A. C. Rencher, "The Popular Critical View of the Isaiah Problem in Light of Statistical Style Analysis," *Computer Studies in the Humanities and Verbal Behavior* 4(1973:149-57; Roger Fowler, "Linguistics, Stylistics: Criticism?" in *Contemporary Essays on Style*, ed. Glen A. Love and Michael Payne (Glenview, Ill.: Scott, Foresman and Company, 1969); C. Douglas Chretien, reviews, *Who Was Junius?* and *A Statistical Method for Determining Authorship: The Junius Letters, 1769-1772* in *Languages* 40 (1964):85-90; Harvey K. McArthur, "KAI Frequency in Greek Letters," *New Testament Studies* 15(1969):339-49; M. Levison, A. Q. Morton, and A. D. Winspear, "The Seventh Letter of Plato," *Mind*, New Series, vol. 77(1968), pp. 109-25; David Wishart and Stephen V. Leach, "A Multivariate Analysis of Platonic Prose Rhythm," *Computer Studies in the Humanities and Verbal Behavior* vol. 3, no. 2(1970):90; S. Michaelson and A. Q. Morton, "Last Words," *New Testament Studies* 8(1972):192-208; W. C. Wake, "Sentence Length Distribution of Greek Authors," *Journal of the Royal Statistical Society* Series A. vol. 70(1957), p. 331; James T. McDonough, Jr., "Computers and the Classics," *Computers and the Humanities* 2(1967):37-40; Noam Chomsky, *Language and Mind* (New York: Harcourt Brace Jovanovich, 1972); Yehuda T. Radday, "The Unity of Isaiah: Computerized Tests in Statistical Linguistics," unpublished reports, Israel Institute of Technology, 1970, pp. 1-172; Claude S. Brinegar, "Mark Twain and the Quintis Curtis Snodgrass Letters: A Statistical Test of Authorship," *Journal of the American Statistical Association* 53(1963):85.

7. A. Q. Morton, *Literary Detection* (New York: Charles Scribner's Sons, 1979).

8. Chretien, reviews, p. 87.

9. See Morton, *Literary Detection,* p. 96.

10. Frederick Mosteller and David L. Wallace, *Inference and Disputed Authorship: The Federalist* (Reading, Mass.: Addison-Wesley, 1964).

11. Morton, pp. 189-91.

12. When Oliver Cowdery transcribed the text of the Book of Mormon as dictated by Joseph Smith, he used very little punctuation. The printer inserted most of the punctuation in the original edition of the Book of Mormon. See B. H. Roberts, *Comprehensive History of The Church of Jesus Christ of Latter-day Saints, Century I,* 6 vols. (Provo, Utah: Brigham Young University Press, 1957), 1:114.

13. These word counts were done using the computerized tapes of the Book of Mormon developed by Elden Ricks and Translation Services of Brigham Young University.

14. Some arbitrary definitions were made. Since, in Mormon theology, the term *Lord* can refer either to God the Father or to his son Jesus, we classified Deity as three distinct authors: the Father, the Lord, and Jesus. We also made the definition that the Lord as quoted by Isaiah is different from Isaiah and also from the Lord in the rest of the book. Our statistical studies showed that these divisions were largely unnecessary.

15. For excerpts from the writings of Joseph Smith, Sidney Rigdon, Parley P. Pratt, Oliver Cowdery, and William W. Phelps, we used a computer disk prepared by Alan J. Phipps (see Phipps, "Lectures on Faith," cited in note 6). We are indebted to Jim Callister for providing this disk. Joseph Smith's writings were taken from articles in the *Messenger and Advocate,* his journal, and letters to various individuals. Joseph Smith's writings included in this study are his own words. This is important, since many works attributed to Joseph Smith were actually written by his scribes or others. See Phipps, "Lectures on Faith," for further information. Sidney Rigdon's writings were taken from the *Evening and Morning Star* and the *Messenger and Advocate.* Parley P. Pratt's works were *A Voice of Warning* and *A Short Account of a Shameful Outrage.* Oliver Cowdery's writings were taken from six letters published in the *Messenger and Advocate.* W. W. Phelps's excerpts were from the *Evening and Morning Star* and the *Messenger and Advocate.* The Doctrine and Covenants sections used in this study were 101 and 104. Solomon Spaulding's writings consisted of five random selections from *Manuscript Found.*

16. We included "The Paracletes," *Times and Seasons,* 6:891-92, 917-18, to determine whether any of our 1830 contemporaries appears to be the author of this unsigned article. Our results were consistently inconsistent—a strong indication that none of the authors used in our study wrote this selection.

17. D. F. Morrison, *Multivariate Statistical Methods* (New York: McGraw-Hill, 1976), chap. 5.

18. Rather than use this frequency, we generally used the arc sine transformation of the frequency for statistical requirements. The program RUMMAGE was used on all MANOVA analyses. See G. R. Bryce, "MAD: An Analysis of Variance Program for Unbalanced Designs," *Journal of the Royal Statistical Society,* Series C (Applied Statistics), vol. 24 (London, 1974), p. 35.

19. The result remained true even when we removed formal words reflecting nineteenth-century religious style from the analyses (*hath, unto,* etc.). The results depend as much on words such as *and, of, for* as on any of the other words.

20. We used a hierarchical clustering algorithm and the Mahalanobis distance function (see P. C. Mahalanobis, "On the Generalized Distance in Statistics," *Proceedings of National Institute of Sciences* 12[India, 1936]:49).

21. See note 16.

22. The discriminant functions can also be used to examine the coefficients of each function so as to possibly identify it as a meaningful new variable. We did not attempt this, but the coefficients are available for someone who may wish to investigate further the nature of the differences among authors.

23. Eighteen discriminant functions were used even though only six were statistically significant. (The two *18's* are coincidental. These numbers will usually be different.)

24. Morton, *Literary Detection,* p. 191.

25. Ibid., pp. 132-37.

26. Adams and Rencher, "The Popular Critical View of the Isaiah Problem," pp. 149-57; Adams and Rencher, "A Computer Analysis of the Isaiah Authorship Problem," pp. 95-102.

27. Morton, *Literary Detection,* p. 185; cf. pp. 186-88.

28. Ibid., pp. 184-85.

29. Ibid., pp. 134-36, 142-43.

30. Alan J. Phipps, "The Lectures on Faith: An Authorship Study," M.A. thesis, Brigham Young University, 1977.