RESEARCH IN THE SCHOOLS
2013, Vol.20, No. 1, i-xi

Editorial: Evidence-based Guidelines for Avoiding Poor Readability in Manuscripts Submitted to Journals for Review for Publication

Anthony J. Onwuegbuzie
Sam Houston State University

Marla H. Mallette
Binghamton University

Eunjin Hwang and **John R. Slate**
Sam Houston State University

In a series of editorials and articles, Onwuegbuzie and colleagues have discussed the utility of writing with discipline when preparing scholarly works for consideration for publication in journals. Moreover, these authors demonstrated the importance of avoiding violations to the American Psychological Association (APA) style guide (APA, 2010) in the abstract and the body of the manuscript, as well as the reference list and table sections of empirical research articles. These editorials have shown that authors are statistically and practically significantly more likely to have their manuscripts rejected for publication if they do not write with discipline. However, in addition to adhering to APA style, this current study provides evidence that important consideration also should be given to the readability of the manuscript. Using the Flesch Reading Ease (RE) and Flesch-Kincaid Grade Level (GL), two common and easily accessible readability formulas, an analysis of the readability of adjudicated manuscripts revealed the following: (a) quantitative research articles had statistically and practically significant lower RE scores and higher Flesch-Kincaid GLs than did qualitative research articles, (b) manuscripts with Flesch RE scores between 0 and 30 were 1.64 more times less likely to be rejected than were manuscripts with Flesch RE scores greater than 30, and (c) manuscripts with Flesch-Kincaid GL scores of 16 and above were 4.55 times less likely to be rejected than were manuscripts with Flesch-Kincaid GL scores less than 16.

Correspondence should be addressed to Anthony J. Onwuegbuzie, Department of Educational Leadership and Counseling, Box 2119, Sam Houston State University, Huntsville, Texas 77341-2119, or E-Mail: tonyonwuegbuzie@aol.com

In a series of editorials and articles, Onwuegbuzie and colleagues have discussed the utility of writing with discipline when preparing scholarly works for consideration for publication in journals. Moreover, these authors demonstrated the importance of avoiding violations to the American Psychological Association (APA) style guide (APA, 2010) in the abstract (Hahs-Vaughn, Onwuegbuzie, Slate, & Frels, 2009) and the body of the manuscript (Onwuegbuzie & Combs, 2009; Onwuegbuzie, Combs, Slate, & Frels, 2010), as well as the reference list (Onwuegbuzie, Combs, Frels, & Slate, 2011; Onwuegbuzie, Frels, & Slate, 2010; Onwuegbuzie & Hwang, 2012, 2013; Onwuegbuzie, Hwang, Combs, & Slate, 2012; Onwuegbuzie, Hwang, Frels, & Slate, 2011; Onwuegbuzie, Waytowich, & Jiao, 2006) and table (Frels, Onwuegbuzie, & Slate, 2010) sections of empirical research articles. These editorials have shown that authors are statistically and practically significantly more likely to have their manuscripts rejected for publication if they do not write with discipline. For example, Onwuegbuzie et al. (2010) documented that manuscripts that contained nine or more different APA errors were 3.00 times (95% confidence interval [CI] = 1.31, 6.87) more likely to be rejected than were manuscripts containing less than nine different APA errors. Further, Onwuegbuzie et al. (2006) reported that authors who made more than three citation errors were approximately four times more likely (odds ratio = 4.01; 95% CI = 1.22, 13.17) to have their manuscripts rejected than were authors who made three or less citation errors. Adhering to APA style in various sections of a manuscript is not the only way to write with discipline. In fact, when writing with discipline, we contend that it is also advantageous for authors to consider the readability of the manuscript.

Readability

Readability is grounded in the idea that the text "becomes readable, most agree, when variables in a text interact with those in a reader to make the writing easy to understand" (Harris & Hodges, 1995, p. 204). As such, the difficulty (or ease) of text is inextricably bound to readers. Thus, in measuring readability, it is essential to examine both the text and the readers' comprehension of the text. However, when derived solely from the text, readability serves as an *estimate*, or *prediction* of the text's difficulty. Readability formulas provide a mechanism for estimating readability by using "counts of language variables in a piece of writing to provide an index of probable difficulty for readers" (Klare, 1974/1975, p. 64).

In the field of literacy, the construct of readability holds a somewhat dubious distinction. That is, for the past century, readability has had a consistent presence in educational research, with more than 1,000 references in the professional literature (Harris & Hodges, 1995). Further, since the development of the first readability formula (Lively & Pressey, 1923), researchers endeavored to determine the features within a text that best predicted difficulty. By the mid-eighties, interest in perfecting readability estimates resulted in the development of more than 100 readability formulas (Klare, 1984); yet, most of these formulas did not endure, with only approximately 20 formulas being used by teachers and researchers (Pearson & Hiebert, 2013). Interestingly, after an examination of the plethora of readability formulas, Klare (1984) determined that, in terms of traditional readability formulas, those formulas with the most predictive power rely on counts of two variables (i.e., syntax and semantics).

Although syntax and semantics as broadly conceptualized constructs seem to represent logical variables in predicting text difficulty, the issue with traditional readability formulas resides in the simplistic operationalizing of these variables. That is, the syntax variable includes counts of length (e.g., sentence length), and the semantics variable involves either counts of word lengths (e.g., number of syllables, number of multisyllabic words) or word frequencies (e.g., number of commonly used words). Nevertheless, readability formulas utilizing these basic counts of words and sentences are indeed highly reliable in predicting text difficulty (see e.g., Graesser, McNamara, & Kulikowich, 2011; Klare, 1984). In addition, Graesser et al. (2011) documented that the relationships among three of the most commonly used formulas: (a) Flesch-Kincaid Grade Level/Reading Ease, (b) Degree of Reading Power, and (c) Lexiles, were exceptionally strong (r = .89 to .94). Although highly reliable, readability formulas, by nature of their design, are limited in purpose and scope. As such, it is essential to be cognizant of their uses, and misuses, along with understanding what information formulas can and cannot provide.

Readability formulas provide an index, a single metric. This index often is expressed as a grade level or a numerical scaled score aligned to a grade level (Graesser et al., 2011). Thus, the index estimates the potential difficulty of a given text in relation to the reader's grade level, with the underlying assumption that grade level captures the requisite skills and knowledge needed by the reader. In contrast, the information gleaned from readability formulas does not define readability, nor explain text difficulty. Perhaps, and most importantly, the text features used in calculating readability (e.g., word counts and sentence length) do not *cause* text difficulty. Using readability formulas in practices that do not adhere to these limitations undermines the standards of correlational research. One prevalent example is in using readability formulas to produce (or to revise) text. This practice is unproductive and often produces the opposite result—less comprehensible text (Davison & Kantor, 1982; Klare, 1974/1975; Pearson & Hiebert, 2013). In examining text in which authors were asked to reduce the readability level, Davison and Kantor (1982) noted that, "the most successful changes in the text often run directly counter to what readability formulas would suggest, and that the most unsuccessful changes are those motivated by the strictures of the readability formulas" (p. 191). Consider, for example, the practice of sentence chopping. If a long sentence is split into two short sentences by removing a complex clause, then the reader is required to infer the connection between the two ideas. Thus, the shorter sentences decrease the readability level; but actually increase difficulty by placing more demands on the reader. In essence, "merely shortening words and sentences to improve readability is like holding a lighted match under a thermometer when you want to make your house warmer" (Klare, 1984, pp. 717-718).

Unlike predicting readability, producing readable text cannot be simplified to a few variables. In fact, it is quite the opposite. Klare's (1984) review of 15 texts on *clear (readable) writing* resulted in a list of 156 different variables that contributed to writing clarity. Within the list of variables, Klare noted that many pairs of ideas were contradictory, and explained, "The problem appeared to be that one pair might apply in a particular set of circumstances and other in another set. But the problem of which to use and when to use it remains unclear" (p. 715). Sawyer (1991) determined the presence of similar inconsistencies in a review of research on how expert revisers revise text. That is, the revisions made by expert revisers improved text; yet, the text improvements were not consistent by reviewer. Also, the expert revisers' explanation of how they revised often did not match what they actually did in their revisions.

Readability is a key variable examined in this current study. Because we considered whether any relationships prevailed between readability and manuscript type, demographics, and dispositions, we believed that it was important to contextualize this variable. Using a readability formula to calculate an index of text difficulty is easy, understanding what makes a text difficult and crafting readable text, are far more complex.

Academic Writing and Readability

Linguistic complexity is characteristic of academic writing. According to Chafe and Danielewicz (1987), the number of words per sentence of academic writers is normally distributed with a mean of 24 words. In contrast, the average length of spoken utterances is 18 words. An additional characteristic of academic writing is difficult text (Gazni, 2011; Metoyer-Duran, 1993). Gazni (2011) examined the relationship between readability estimates (i.e., Flesch Reading Ease Scores of abstracts only) and citation rates for articles published between 2000 and 2009 from the five institutions (e.g., Harvard) that receive the largest number of citations. Based on an analysis of approximately 260,000 abstracts, spanning 22 disciplines, Gazni established a statistically significant, negative relationship between text difficulty and citation rates; that is, the more difficult the text, the more it was cited. Further, the Flesch Reading Ease Scores, which ranged from an average of 12.6 in pharmacology/toxicology to an average of 25.6 in mathematics, all fell into the *Very Difficult* range or college grade level (i.e., 1-30), the highest level of text difficulty.

Somewhat similar to our own investigation, Metoyer-Duran (1993) examined whether readability estimates differed significantly among published, accepted, and rejected manuscripts and abstracts from *College and Research Libraries* during the 1990-1991 period. Metoyer-Duran determined that the readability estimates of manuscripts accepted for publication were significantly different from the readability estimates of manuscripts rejected for publication. For example, the mean Flesch Reading Ease Score was 28.04 for accepted manuscripts and 30.77 for rejected manuscripts. Thus, manuscripts that were accepted for publication contained text that was more difficult than was the text in manuscripts that were rejected.

This current study extends the findings of Metoyer-Duran (1993) in three ways. First, the current investigation involves examining the extent to which readability estimates predict whether or not manuscripts are rejected and/or accepted. Second, this study involves determining whether there are any significant relationships between readability estimates and manuscript types and/or demographics. Third, in this study, readability estimates from the Flesch Reading Ease Score and Flesch–Kincaid Grade Level are determined for the entire manuscript, whereas Metoyer-Duran (1993) used only five paragraphs from each manuscript.

The Flesch Reading Ease and the Flesch–Kincaid Grade Level are so common that word processing programs provide these indexes for any text in less than a few seconds. Yet, both are quite easy to calculate, which we explain in the following sections.

Flesch Reading Ease

The formula for calculating the Flesch Reading Ease Score is

$$206.835 - 1.015\left(\frac{\text{total words}}{\text{total sentences}}\right) - 84.6\left(\frac{\text{total syllables}}{\text{total words}}\right)$$

For example, the Flesch Reading Ease Score for the last paragraph in the *Readability* section of this editorial can be calculated as follows:

$$206.835 - 1.015\left(\frac{64}{3}\right) - 84.6\left(\frac{136}{64}\right)$$

This calculation yields a rounded Flesch Reading Ease Score of 5.4. According to Flesch (1946), the highest (i.e., easiest) Flesch Reading Ease Score is approximately 100, with a Flesch Reading Ease Score between 90 and 100 representing text that potentially can be understood by fifth-grade students; a Flesch Reading Ease Score between 60 and 70 representing text that potentially can be understood by eighth- to ninth-grade students; and a Flesch Reading Ease Score between 0 and 30 representing text that potentially can be understood by college graduate students (Flesch, 1946). Theoretically, the Flesch Reading Ease Score does not have a lower bound because it is possible to make the score as low as desired by deliberately including words with many syllables. Using the aforementioned criteria, it can be assumed that the Flesch Reading Ease Score of the last paragraph in the *Readability* section of 5.4 very much represents text that is appropriate for many of the readership of *Research in the Schools*.

Flesch–Kincaid Grade Level

The formula for the Flesch-Kincaid Grade Level essentially converts the reading ease score to a U.S. grade level in order for educators, parents, students, and others to assess the readability level of text via a common index. This grade level, in turn, can be converted to the number of years of education broadly required to understand the underlying text when the formula yields a grade level that exceeds 10. Specifically, the formula for the Flesch-Kincaid Grade Level is

$$0.39\left(\frac{\text{total words}}{\text{total sentences}}\right) + 11.8\left(\frac{\text{total syllables}}{\text{total words}}\right) - 15.59$$

Interestingly, multisyllabic words affect this score significantly more than the Flesch Reading Ease score. Returning to our example of the last paragraph in the *Readability* section of this editorial, Microsoft Office Word calculates the Flesch-Kincaid Grade Level also as 17.8.

The Flesch Reading Ease Score, and the Flesch–Kincaid Grade Level are available in Microsoft Office Word's *readability statistics*, along with average sentence length, average number of sentences per paragraph, average number of characters per word, and the percentage of passive-voice sentences. In order to enable Microsoft Office Word's 2013 readability program, the user should click on *FILE*, then *Options*, then *Proofing*, and then check *Show readability statistics*. These steps will lead to readability statistics being displayed whenever the user conducts a spell-check of her/his document. (However, it should be noted that, at the time of writing, the Mac version of Microsoft Office Word of the Flesch–Kincaid Grade Level has been artificially capped at grade-level 12.)

Purpose of Study

The purpose of our study is to examine the importance of readability on the quality of empirical manuscripts submitted to a journal for consideration for publication, as indicated by their eventual disposition (i.e., accept/revise and resubmit vs. reject). Specifically, the following five research questions were addressed:

1. What is the distribution of readability scores among manuscripts submitted to a journal?
2. What is the relationship among readability scores among manuscripts submitted to a journal?
3. What is the relationship between readability scores and style guide errors (i.e., citation errors) among manuscripts submitted to a journal?
4. What is the relationship between readability scores and select demographic characteristics (i.e., number of authors, genre of manuscript) among manuscripts submitted to a journal?
5. What is the relationship between readability scores and manuscript disposition among manuscripts submitted to a journal?

Method

Sample Size and Procedures

To analyze the distribution and predictability of readability estimates among manuscripts submitted to a journal, we examined 63 manuscripts submitted to *RITS* over a 3-year period (i.e., 2011-2014). These manuscripts represented approximately 40% of all manuscripts submitted to this journal over this time frame, which made these findings generalizable at the very least to the population of manuscripts submitted to *RITS*. The sample size of 63 was selected via an a priori statistical power analysis. Specifically, it represented the sample size needed to detect a moderate multivariate relationship (i.e., discriminant analysis; f = .25) simultaneously for six dependent measures (i.e., six readability scores) between the two groups (i.e., manuscripts that were rejected vs. manuscripts that were not rejected) at the 5% level of statistical significance and a power of .80.

One of the editors of the *Research in the Schools* (i.e., lead author of this editorial) meticulously documented every citation error committed by these 63 sets of authors. Each manuscript took up to 2 hours to identify all the citation errors—representing as much as 126 hours of coding. In addition, the editor noted several demographic features of the manuscript (e.g., number of authors, gender of first author, genre of manuscript), as well as the disposition of the manuscript. Finally, using Microsoft Office Word's *Show readability statistics* option, the editor computed the following 10 readability statistics: number of words, number of characters, number of paragraphs, number of sentences, average sentence length, average number of sentences per paragraph, average number of characters per word, the percentage of passive-voice sentences, Flesch Reading Ease, and the Flesch–Kincaid Grade Level. As such, the data set created by these editors was extremely rich, representing a data set that only journal editors have the opportunity to develop.

Analysis

Research Question 1. What is the distribution of readability scores among manuscripts submitted to a journal? Descriptive statistics (i.e., measures of central tendency, measures of dispersion, measures of distributional shape) were used to address this research question. With regard to measures of central tendency and measures of dispersion, means and standard deviations, respectively, were computed for each of the readability measures. With regard to measures of distributional shape, skewness and kurtosis coefficients were computed to help assess the (multivariate) normality of the each measure. To obtain the readability estimates, for each manuscript, only the abstract and body of the manuscript were assessed. That is, the reference list, tables, and figures were excluded from the readability analysis.

Research Question 2. What is the relationship among readability scores among manuscripts submitted to a journal? To address this research question, the following six primary readability indices were involved: average number of sentences per paragraph, average number of words per sentences, average number of characters per word, the percentage of passive-voice sentences, Flesch Reading Ease, and the Flesch–Kincaid Grade Level. Depending on the extent to which the normality assumption held, either a parametric correlation (i.e., Pearson *r*) or nonparametric correlation (i.e., Spearman's rho) was used to assess these intercorrelations. Effect sizes were interpreted for all statistically significant findings (Cohen, 1988).

Research Question 3. What is the relationship between readability scores and style guide errors (i.e., citation errors) among manuscripts submitted to a journal? The citation error was selected as a criterion variable because it represents by far the most common APA error, with between 88.6% (Onwuegbuzie, Combs, et al., 2011) and 91.8% (Onwuegbuzie, Frels, et al., 2010) of authors committing one or more citation errors. To address this research question, the same six primary readability indices were involved. Depending on the extent to which the normality assumption held, either a series of parametric correlations (i.e., Pearson *r*) or nonparametric correlations (i.e., Spearman's rho) was used to assess the relationships between readability scores and the number of citation errors. In addition, depending on the extent to which the normality assumption held, a multiple regression analysis was used to determine which readability variables predicted the number of citation errors.

Research Question 4. What is the relationship between readability scores and select demographic characteristics (i.e., number of authors, genre of manuscript) among manuscripts submitted to a journal? When examining the number of authors, the six primary readability indices were involved. However, when examining the genre of manuscript, the Flesch Reading Ease and Flesch-Kincaid Grade Level were the only readability indices involved. Depending on the scale of measurement underlying the selected demographic variable (i.e., ratio vs. nominal), either a parametric/nonparametric correlation (i.e., number of authors; ratio scale; dichotomous data), or a parametric/nonparametric analysis of variance (i.e., genre of manuscript nominal scale; polytomous data) was used.

Research Question 5. What is the relationship between readability scores and manuscript disposition among manuscripts submitted to a journal? To avoid the possibility of multicollinearity (i.e., both the Flesch Reading Ease scores and Flesch-Kincaid Grade Levels are functions of the average number of words per sentence), the Flesch Reading Ease and Flesch-Kincaid Grade Level were the only readability indices involved. Depending on the extent to which the multivariate normality assumption held, either a discriminant analysis or logistic regression was used to determine, which readability variables as a set, if any, discriminated the two sets of manuscripts (i.e., manuscripts that were rejected vs. manuscripts that were not rejected).

Results

Research Question 1. **What is the distribution of readability scores among manuscripts submitted to a journal?**

Table 1 presents the means, standard deviations, minimum values, and maximum values of the 10 readability indices. Most notable is the fact that the number of sentences per paragraph ranged from 3.1 sentences to 49.0 sentences. According to the authors of APA (2010, p. 68), paragraphs should not be longer than one double-spaced manuscript page. Thus, not only are paragraphs that contain as many as 49 sentences likely to be unwieldy, but they violate APA stipulations. However, it should be noted that only one manuscript contained 49-sentence paragraphs, on average. Perhaps not surprisingly, this manuscript was rejected. For the remaining manuscripts, the mean number of sentences per paragraph ranged from 3.1 to 7.4.

Of particular interest was the fact that the mean number of words per sentence was 23.39. This figure is very consistent with the mean number of 24 identified by Chafe and Danielewicz (1987) for academic writers. Also, in comparison to number of words per sentence in the abstracts of journal articles from the top five most cited universities (Gazni, 2011), this number is within the range of all journals across disciplines (i.e., 24.2-29.0), and slightly below the grand mean of 26.2. In addition, word length was quite comparable between the two studies with the mean word length of 5.8 in Gazni's (2011) study and 5.41 in the present study.

Another notable finding was that the percentage of passive sentences ranged from 5% to 43%, with a mean of 21.52%. Interestingly, 12.5% of manuscripts contained sentences for which at least one third were passive, 40.6% of manuscripts contained sentences for which at least one fourth were passive, 51.6% of manuscripts contained sentences for which at least one fifth were passive, and 93.7% of manuscripts contained sentences for which at least 10% were passive.

RESEARCH IN THE SCHOOLS
2013, Vol.20, No. 1, i-xi

Table 1

Descriptive Statistics for the Readability Measures

Readability Measure	*M*	*SD*	Minimum	Maximum
Number of words	6,060.72	2,569.27	1,901	13,595
Number of characters	35,428.47	16,987.46	10,291	118,048
Number of paragraphs	95.08	84.87	21	533
Number of sentences	247.28	109.44	85	718
Average sentence per paragraph	5.42	5.60	3.1	49.0
Average word per sentence	23.39	3.14	12.2	30.2
Average characters per word	5.41	0.26	4.7	5.9
Percentage of passive sentences	21.52	8.68	5.0	43.0
Flesch Reading Ease	29.80	9.08	12.10	61.10
Flesch-Kincaid Grade Level	14.21	1.59	8.80	17.70

Most interesting were the descriptive statistics pertaining to Flesch Reading Ease and Flesch-Kincaid Grade Level. With regard to Flesch Reading Ease, the mean score was 29.8, which represents text that potentially can be understood by university graduates. The Flesch Reading Ease scores ranged from 12.1 (i.e., university graduate level) to 61.1 (i.e., 8th- to 9th-grade level). Slightly more than one half (i.e., 53.1%) of the manuscripts yielded Flesch Reading Ease scores between 12.1 and 30.0. With respect to Flesch-Kincaid Grade Levels, the grade levels ranged from 8.8 to 17.7, with a mean of 14.21. Interestingly, exactly one half (i.e., 50.0%) of the manuscripts yielded a Flesch-Kincaid Grade Level of 14.0.

Using Onwuegbuzie and Daniel's (2002) criteria for a standardized skewness coefficient (i.e., skewness coefficient divided by its standard error) and a standardized kurtosis coefficient (i.e., kurtosis coefficient divided by its standard error), wherein standardized skewness coefficients and standardized kurtosis coefficients that lie outside the ±3 range indicate serious departures from normality, of the six primary readability indices, only one of them suggested non-normality, namely, the average sentence per paragraph. Interestingly, this non-normality was primarily caused by the one manuscript that contained 49-sentence paragraphs, on average. Most importantly, the Flesch Reading Ease Scores and the Flesch-Kincaid Grade Levels were normally distributed—consistent with the findings of Chafe and Danielewicz (1987).

Research Question 2. **What is the relationship among readability scores among manuscripts submitted to a journal?**

Because five of the six primary readability indices were normally distributed, a series of Pearson *r*s was used to determine their intercorrelations. Table 2 presents these intercorrelations. It can be seen from this table that, after applying the Bonferroni adjustment to control for the inflation of Type 1 error, the following four correlations were statistically significant: (a) between Flesch Reading Ease and the average number of characters per word; and between Flesch-Kincaid Grade Level and (b) the average number of words per sentence, (c) the average number of characters per word, and (d) Flesch Reading Ease. Using Cohen's (1988) criteria, and consistent with the literature, the correlation between Flesch Reading Ease and Flesch-Kincaid Grade Level (i.e., $r = -.89$) represented a very large inverse relationship (Kincaid, Fishburne, Rogers, & Chissom, 1975).

Table 2

Intercorrelations Among Primary Readability Measures (n = 63)

Readability Measure	2	3	4	5	6
1. Average sentence per paragraph	.18	-.13	-.08	-.09	-.02
2. Average word per sentence		-.08	.27	-.28	.63*
3. Average characters per word			-.08	-.83*	.61*
4. Percentage of passive sentences				-.08	.21
5. Flesch Reading Ease					-.89*
6. Flesch-Kincaid Grade Level					

* Statistically significant at the Bonferroni-adjusted value of .0033 (i.e., .05/15)

Research Question 3. What is the relationship between readability scores and style guide errors (i.e., citation errors) among manuscripts submitted to a journal?

A series of Pearson *r*s revealed no statistically significant relationship between the number of citation errors committed by the author and all six primary readability indices, namely: average number of sentences per paragraph (r = -.05, p = .70), average number of words per sentence (r = .04, p = .75), average number of characters per word (r = -.07, p = .59), the percentage of passive-voice sentences (r = -.05, p = .72), Flesch Reading Ease (r = .12, p = .38), and the Flesch–Kincaid Grade Level (r = -.10, p = .44). Thus, it should not be surprising that a multiple regression analysis containing all six primary readability measures in the model was not statistically significant (F[6, 56] = 0.56, p = .76), only explaining 5.8% of the variance. Further, none of the readability indices made a statistically significant contribution to the model (i.e., non-significant standardized and structured coefficients) Thus, these findings indicate that the readability of manuscripts represents a very distinct construct from citation errors. As such, authors who write manuscripts with readability that are non-optimal are not necessarily those authors who commit the most common APA error, namely, the citation error (Onwuegbuzie, Combs, et al., 2011; Onwuegbuzie, Frels, et al., 2010).

Research Question 4. What is the relationship between readability scores and select demographic characteristics (i.e., number of authors, genre of manuscript) among manuscripts submitted to a journal?

After applying the Bonferroni adjustment to control for the inflation of Type I error (i.e., adjusted α = .05/6 = .00833), a series of Pearson *r*s revealed no statistically significant relationship (i.e., p > .0083) between the number of manuscript pages and the following five primary readability indices: average number of sentences per paragraph (r = .16), average number of characters per word (r = .19), the percentage of passive-voice sentences (r = .19), Flesch Reading Ease (r = .04), and the Flesch–Kincaid Grade Level (r = -.26). However, a statistically significant relationship emerged between the number of authors of a manuscript and the average number of words per sentence (r = -.55, p < .001), which, using Cohen's (1988) criteria, represented a large effect size. This correlation indicated that manuscripts that involved more authors tended to contain a smaller average number of words per sentence.

After applying the Bonferroni adjustment to control for the inflation of Type I error (i.e., adjusted α = .05/2 = .025), the first analysis of variance (ANOVA) revealed a statistically significant difference in Flesch Reading Ease scores among the quantitative, qualitative, and mixed research manuscripts, F(2, 60) = 4.12, p = .022. A Scheffé post-hoc test revealed no statistically significant difference in Flesch Reading Ease scores between the mixed research manuscripts (M = 30.84, SD = 8.00) and the qualitative research manuscripts (M = 35.05, SD = 10.96). However, although a statistically significant difference was not present in Flesch Reading Ease scores between the quantitative research manuscripts (M = 26.79, SD = 8.23) and the mixed research manuscripts, the quantitative research manuscripts had statistically significantly lower Flesch Reading Ease scores than did the qualitative research manuscripts. Cohen's (1988) d effect size associated with this difference between the quantitative and qualitative research manuscripts was large at 0.89. Also interesting was the finding that whereas, for the quantitative research manuscripts, on average, the Flesch Reading Ease scores lay in the 0-30 range (i.e., representing text that potentially can be understood by college graduate students), for both the mixed research manuscripts and the qualitative research manuscripts, on average, the Flesch Reading Ease scores lay outside the 0-30 range.

Similarly, the second ANOVA revealed a statistically significant difference in Flesch-Kincaid Grade Levels among the quantitative, qualitative, and mixed research manuscripts, F(2, 60) = 4.23, p = .020. A Scheffé post-hoc test did not reveal a statistically significant difference in Flesch-Kincaid Grade Levels between the mixed research manuscripts (M = 14.03, SD = 1.27) and the qualitative research manuscripts (M = 13.23, SD = 1.87). However, although a statistically significant difference was not present in Flesch-Kincaid Grade Levels between the quantitative research manuscripts (M = 14.69, SD = 1.50) and the mixed research manuscripts, the quantitative research manuscripts had statistically significantly higher Flesch-Kincaid Grade Levels than did the qualitative research manuscripts. Cohen's (1988) d effect size associated with this difference between quantitative and qualitative research manuscripts was large at 0.89.

Research Question 5. What is the relationship between readability scores and manuscript disposition among manuscripts submitted to a journal?

A canonical discriminant analyses was conducted to determine which of the Flesch Reading Ease and Flesch-Kincaid Grade Level indices best predicted whether the editor‘s decision for a manuscript was reject versus non-reject (i.e., revise and resubmit or accept). Prior to conducting this analysis, Box's M test was conducted to assess the homogeneity of the variance-covariance matrix involving the four variables of interest (Tabachnick & Fidell, 2007). Box's M statistic was 1.06, which suggested homogeneity of the variance-covariance matrix (F [3, 1476717.97] = 0.34, p = .80), thereby justifying the discriminant analysis. The canonical discriminant analysis revealed a statistically significant canonical function ($\times^2$[2] = 8.83, p = .012; Wilks's Lambda = 0.86). The corresponding canonical correlation was .37, which suggested a medium effect size (Cohen, 1988). In addition, the group centroid (the average score on the discriminant function for students in both groups) for this function was -.37 for manuscripts that were rejected and .42 for manuscripts that were not rejected. These statistics indicated that the discriminant function maximally separated manuscripts that were rejected from manuscripts that were not rejected.

An examination of the standardized canonical discriminant function coefficients (Table 3) revealed that, using a cutoff loading of 0.3 (Lambert & Durand, 1975; Tabachnick & Fidell, 2007), both Flesch Reading Ease and Flesch-Kincaid Grade Level were practically significant. Further, the structure coefficients (i.e., structure matrix) between the independent variable set and the standardized canonical discriminant function (Table 3) indicated that, using a cutoff loading of 0.3 (Lambert & Durand, 1975; Tabachnick & Fidell, 2007), both Flesch Reading Ease and Flesch-Kincaid Grade Level discriminated manuscripts that were rejected and manuscripts that were not rejected. Flesch-Kincaid Grade Level was the most significant predictor of manuscript disposition. The variable with a positive coefficient (i.e., Flesch Reading Ease) suggests that manuscripts high on this measure were more likely to be rejected. Conversely, the measure with a negative coefficient (i.e., Flesch-Kincaid Grade Level) indicates that manuscripts low on this measure were more likely to be rejected.

Table 3

Standardized and Structure Coefficients for Selected Readability Variables

Variable	Standardized Coefficient	Structure Coefficient
Flesch Reading Ease	.79*	.90*
Flesch-Kincaid Grade Level	-.91*	-.99*

* coefficients with effect sizes larger than .3 (Lambert & Durand, 1975)

The next goal was to obtain cut-points for Flesch Reading Ease and Flesch-Kincaid Grade Level that discriminated manuscripts that were rejected from manuscripts that were not rejected. With regard to Flesch Reading Ease, conducting a series of chi-square analyses revealed that manuscripts with Flesch Reading Ease scores between 0 and 30 were statistically significant less likely to be rejected than were manuscripts with higher Flesch Reading Ease scores. More specifically, manuscripts with Flesch Reading Ease scores between 0 and 30 were 1.64 times less likely to be rejected than were manuscripts with Flesch Reading Ease scores greater than 30. Interestingly, after assessing several other cut-points, once the range of Flesch Reading Ease scores included 31, it no longer statistical significantly discriminated manuscripts that were rejected from manuscripts that were not rejected. This result means that 30 was the cut-point for discriminating manuscripts that were rejected from those manuscripts that were not rejected.

Interestingly, the score of 30, the end number in the range of the most difficult text, represents the cut-point not only in this study, but also was noteworthy in the research of Gazni (2011) and Metoyer-Duran (1993). That is, Gazni determined Flesch Reading Ease scores from the abstracts of articles that spanned 22 disciplines published by authors across the five institutions that received the greatest number of citations. The Flesch Reading Ease scores ranged from 12.6 to 25.6, with individual mean scores for each discipline at each institution ranging from 12 to 30. Thus, among the 110 Flesch Reading Ease score means, there was not a single mean score greater than 30. Similarly, Metoyer-Duran (1993) documented that the mean scores of manuscripts rejected for publication in *College and Research Libraries* was greater than 30 (i.e., 30.77), and the mean score for manuscripts accepted for publication was less than 30 (i.e., 28.04).

Similarly, with respect to the Flesch-Kincaid Grade Level scores, the cut-point was Grade 16. That is, manuscripts with Flesch-Kincaid Grade Level scores of 16 and above were statistically significantly less likely to be rejected. Moreover, manuscripts with Flesch-Kincaid Grade Level scores of 16 and above were 4.55 times less likely to be rejected than were manuscripts with Flesch-Kincaid Grade Level scores less than 16. Once the Grade Level score of 15 was included (i.e., Grade Level scores of 15 and above vs. Grade Level scores below 15), it no longer statistically significantly discriminated manuscripts that were rejected from manuscripts that were not rejected.

These two sets of findings allow us to provide a noteworthy evidence-based conclusion. Specifically, from these findings, we conclude that manuscripts submitted to *Research in the Schools* optimally should have a Flesch Reading Ease score between 0 and 30 and a Flesch-Kincaid Grade Level of 16 or higher. Thus, manuscripts with readability scores that fall outside these ranges likely are at a higher risk for rejection. Interestingly, the complete body of our current editorial yielded a Flesch Reading Ease score of 19.6 and a Flesch-Kincaid Grade Level of 16.2. As such, the readability scores of our editorial falls within these ranges, suggesting that it has been written at the appropriate scholarly level.

Discussion of Findings

The study underlying the present editorial is unique in at least five ways. First, it represents the only study wherein the whole bodies (i.e., not including reference lists, tables, and figures) of manuscripts submitted to journals have been examined for readability scores. Second, it likely represents only the second formal attempt to examine the relationship among readability scores among manuscripts submitted to a journal. Third, this editorial represents the first work to examine the link between readability scores and style guide errors—namely, citation errors—among manuscripts submitted to a journal. Fourth, this editorial represents the first study to investigate the relationship between readability scores and select demographic characteristics (i.e., number of authors, genre of manuscript) among manuscripts submitted to a journal. Fifth, and most importantly, this study represents the first attempt to examine the relationship between readability scores from the whole bodies of manuscripts and manuscript disposition among manuscripts submitted to a journal.

The results from this current study demonstrate three unique findings in relation to readability indices and manuscript preparation and submission. Specifically, readability indices differ in relation to the following: (a) number of authors, (b) research tradition, and (c) adjudication decisions. In the section that follows, we will discuss these findings; however, we first remind readers of an important caveat to consider relative to these findings. That is, it is essential to bear in mind that the variables in the readability formulas (i.e., word length and sentence length) likely serve as proxies for a host of other variables. Manuscript preparation and adjudication, like text comprehension, are complex processes that cannot be reduced to simplistic language counts. Literacy scholars have theorized that the language variables in readability formulas capture the following premise regarding text difficulty:

> … it is the complexity of the ideas they are trying to communicate that drives writers to craft prose that ends up being grammatically and semantically more complex—that is, expressed in longer, more complex, and likely more obscure, words and sentences. In short, there is a reason for complexity—the ideas are difficult! (Pearson & Hiebert, 2013, p. 3)

Thus, the difficult and complex content in academic research seems most successfully articulated with sophisticated prose.

Interestingly, though, this sophisticated prose is captured in shorter sentences when written by larger author teams. That is, larger author teams tend to produce manuscripts with fewer words per sentence than do smaller author teams. However, these differences in sentence length, although statistically significant, do not change the Flesch Reading Ease and/or the Flesch-Kincaid Grade Level scores. Perhaps, publishable manuscripts crafted by large author teams reflect a collaborative process of writing in which the complexity of academic scholarship is captured more efficiently.

Another difference depicted by readability estimates was related to research methodology. The text in manuscripts in which authors reported on studies using quantitative methods was statistically significantly more difficult than was the text in manuscripts in which authors reported on qualitative studies. Although at first this finding appears counterintuitive, upon deeper reflection, it seems quite appropriate. Qualitative research is predominantly narrative, and as such, written reports are more verbose than that of quantitative research. However, an important consideration is that *more words* does not necessarily equate to more multisyllabic words or more words per sentence. Rather, *more words* quite simply means more text, or more pages of text. Thus, although reporting on qualitative research often requires more text than does reporting on quantitative research, the text itself is less difficult. In considering, for example, the results sections within these two types of manuscripts, the differences in difficulty are quite predictable. Qualitative studies describe phenomenon by capturing the perspectives of the participants. This emic, or insider’s lens, ought to reflect the more common words within spoken, everyday language. In contrast, the results of quantitative research studies detail statistical techniques, which require specialized, technical vocabulary to explain and to describe.

Clearly, the most perplexing finding is that this specialized, technical, difficult, and dense text draws more favorable reviews, and is more likely to be accepted for publication, than is text with lower readability estimates. Although purely speculative, as editors and reviewers, we contend that a plausible explanation resides in the value of and regard for scholarship. Research in education, that truly makes a difference, is a means to: empower participants, improve the quality of education, improve the experiences for future participants, move the field forward, test theories, create theories, and find answers to difficult questions. Goals as lofty as these goals require the production of rigorous, thoughtful, and complex research studies. Further, when authors disseminate their work, by crafting text that is equally thoughtful and rigorous, it is well received by consumers.

Conclusions

Our current editorial provides compelling evidence that readability is indeed an important characteristic of writing with discipline. However, just as citation errors do not cause manuscript adjudication decisions, readability indices do not cause these decisions either. In juxtaposing the findings of this study with the findings from Gazni (2011) and Metoyer-Duran (1993), successful academic writing includes highly sophisticated prose. Therefore, we encourage authors to use readability indices as a gauge in the writing process. Macdonald-Ross (1978) explained how to use readability scores as a *control device*, or a mechanism to provide feedback to writers. Macdonald-Ross highlights the process as follows:

> Providing there is some knowledge of the target audience, the formula can be the engine of a writing-rewriting cycle:
>
> Write → Apply formula → Revise → Apply formula
>
> In this process, the inexactness of the matching process must always be remembered; we should not aim for ultraprecise fits, but we should aim to eliminate gross misfits. (p. 236)

References

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Chafe, W., & Danielewicz, J. (1987). Properties of spoken and written language. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending oral and written language* (pp. 83-113). New York, NY: Emerald Group.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Davison A., & Kantor, R. N. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly, 17*, 187-208.

Flesch, R. (1946). *The art of plain talk*. New York, NY: Harper & Row.

Frels, R. K., Onwuegbuzie, A. J., & Slate, J. R. (2010). Editorial: A step-by-step guide for creating tables. *Research in the Schools*, *17*(2), xxxviii-lix. Retrieved from http://www.msera.org/download/RITS_17_2_Tables.pdf

Gazni, A. (2011). Are the abstracts of high impact articles more readable? Investigating the evidence from top research institutions in the word. *Journal of Information Science, 37*, 273-281. doi:10.1177/0165551511401658

Graesser, A. C., McNamara, D. S., & Kulikowich, J. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*, 223-234. doi:10.3102/0013189X11413260

Hahs-Vaughn, D. L., Onwuegbuzie, A. J., Slate, J. R., & Frels, R. K. (2009). Editorial: Bridging research-to-practice: Enhancing knowledge through abstracts. *Research in the Schools*, *16*(2), xxxvii-xlv. Retrieved from http://www.msera.org/download/RITS_16_2_Abstracts.pdf

Harris, T. L., & Hodges, R. E. (Eds.). (1995). *The literacy dictionary: The vocabulary of reading and writing.* Newark, DE: The International Reading Association.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel.* Research Branch Report 8-75. Millington, TN: Naval Air Station Memphis.

Klare, G. R. (1974/1975). Assessing readability. *Reading Research Quarterly, 10*, 62-102. doi:10.2307/747086

Klare, G. R. (1984). Readability. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (Vol. 1, pp. 681-744). New York, NY: Longman.

Lambert, Z. V., & Durand, R. M. (1975). Some precautions in using canonical analysis. *Journal of Market Research*, *12*, 468-475. doi:10.2307/3151100

Lively, B., & Pressey, S. (1923). A method for measuring the “vocabulary burden” of textbooks. *Educational Administration and Supervision, 99*, 389-398.

Macdonald-Ross, M. (1978). Language in texts. *Review of Research in Education, 6,* 229-275. doi:10.2307/1167247

Metoyer-Duran, C. (1993). The readability of published, accepted, and rejected papers appearing in *College & Research Libraries. College & Research Libraries, 54,* 517-526.

Onwuegbuzie, A. J., & Combs, J. P. (2009). Writing with discipline: A call for avoiding APA style guide errors in manuscript preparation. *School Leadership Review*, *4*, 116-149.

Onwuegbuzie, A. J., Combs, J. P., Frels, R. K., & Slate, J. R. (2011). Editorial: Citation errors revisited: The case for *Educational Researcher*. *Research in the Schools, 18*(1), i-xxxv. Retrieved from http://www.msera.org/download/RITS_18_1_Complete.pdf

Onwuegbuzie, A. J., Combs, J. P., Slate, J. R., & Frels, R. K. (2010). Editorial: Evidence-based guidelines for avoiding the most common APA errors in journal article submissions. *Research in the Schools, 16*(2), ix-xxxvi. Retrieved from http://msera.org/download/RITS_16_2_APAErrors6th.pdf

Onwuegbuzie, A. J., & Daniel, L. G. (2002). Uses and misuses of the correlation coefficient. *Research in the Schools, 9*(1), 73-90.

Onwuegbuzie, A. J., Frels, R. K., & Slate, J. R. (2010). Editorial: Evidence-based guidelines for avoiding the most prevalent and serious APA error in journal article submissions—The citation error. *Research in the Schools*, *17*(2), i-xxiv. Retrieved from http://www.msera.org/download/RITS_17_2_Citations.pdf

Onwuegbuzie, A. J., & Hwang, E. (2012). *Reference list errors in manuscripts submitted to a journal for review for publication*. Unpublished manuscript, Sam Houston State University, Huntsville, TX.

Onwuegbuzie, A. J., & Hwang, E. (2013). Reference list errors in manuscripts submitted to a journal for review for publication. *International Journal of Education,* *5*(1), 1-14. doi:10.5296/ije.v5i2.2191. Retrieved from http://www.macrothink.org/journal/index.php/ije/article/view/2191

Onwuegbuzie, A. J., Hwang, E., Combs, J. C., & Slate, J. R. (2012). Editorial: Evidence-based guidelines for avoiding reference list errors in manuscripts submitted to journals for review for publication: A replication case study of *Educational Researcher*. *Research in the Schools, 19*(2), i-xvi.

Onwuegbuzie, A. J., Hwang, E., Frels, R. K., & Slate, J. R. (2011). Editorial: Evidence-based guidelines for avoiding reference list errors in manuscripts submitted to journals for review for publication. *Research in the Schools, 18*(2), i-xli. Retrieved from http://msera.org/rits.htm

Onwuegbuzie, A. J., Waytowich, V. L., & Jiao, Q. G. (2006, December). Bibliographic errors in articles submitted to scholarly journals: The case for *Research in the Schools*. *Academic Exchange Extra*. Retrieved from http://www.unco.edu/AE-Extra/2006/12/ Jiao.html

Pearson, P. D., & Hiebert, E. H. (2013). *The state of the field: Qualitative analysis of text complexity*. Reading Research Report #13.01. Text Project, Inc. Santa Cruz, CA. Retrieved from http://textproject.org/research/

Sawyer, M. (1991). A review of research in revising instructional text. *Journal of Reading Behavior, 23*, 307-333. doi:10.1080/10862969109547744

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon. Retrieved from http://www.census.gov/hhes/www/income/data/statemedian/