

Reading Difficulty Levels of Selected Articles in the *Journal of Research in Music Education* and *Journal of Historical Research in Music Education*

Alexandra H. Humphreys
Arizona State University—Phoenix, U.S.A.

Jere T. Humphreys
Arizona State University—Tempe, U.S.A.

Abstract

Readability formulas are used widely in education, and increasingly in business and government. Over 30 years of research on more than 200 readability formulas has demonstrated moderate to strong predictive correlations with reading comprehension. In this study, five well-known readability formulas correlated highly with each other when applied to selected recent historical articles (N = 22) from two music education research journals. The mean level of difficulty (readability) for all 22 articles was grade 14.04, near the beginning of the second year of college. Since research shows that most people read below their highest completed school grade and also prefer easier materials, this is probably an appropriate level of difficulty for the presumptive readers of these two journals (i.e., holders of undergraduate and graduate degrees). Professors, librarians, and others responsible for guiding students toward reading material at appropriate levels of readability could benefit from these results.

Scholars have analyzed various aspects of the research literature in music education. Butler (1973) examined music psychology literature from the 19th century, while others have focused on research papers presented at selected conferences, with emphasis on such variables as numbers of papers presented and demographic characteristics of the authors (Cooper & Bayless, 2008; Hedden, 1992, 1993). Kantorski (1995) reported on the content of doctoral dissertations related to string music education. Humphreys, Bess, and Bergee (1997) and Preston and Humphreys (2007)¹ examined topics and selected demographic variables related to the production of dissertations on the history of music education from the 1920s through the 1990s.

¹ See these two studies for reviews of the literature on master's theses and doctoral dissertations in music education.

Other researchers have examined scholarly journals related to music education. There are published studies on the composition of the editorial committees of the *Journal of Research in Music Education* (JRME) (Humphreys & Stauffer, 2000) and *Bulletin of Historical Research in Music Education* (BHRME) (Humphreys, 1999), now called the *Journal of Historical Research in Music Education* (JHRME). There are also studies on research methodologies and topics of articles in the BHRME (McCarthy, 1999), JRME (Yarbrough, 1984, 2002), *Contributions to Music Education* (Hall, 1998), and *Bulletin of the Council for Research in Music Education* (Stabler, 1986). Characteristics of subjects (Kratus, 1992) and research samples (Ebie, 2002) employed in studies reported in selected music education research journals have also been investigated. Finally, a body of studies has shed light on the authors and citations of articles in selected research journals in music education (see Latimer, 2011).

Readability Research

The first known readability formula was published in the United States in 1923, following attempts in Germany, Russia, and elsewhere in the late 19th century to match students with materials at appropriate levels of difficulty (Lively & Pressey, 1923). In 1949, the developer of two of the best known readability formulas, Rudolf Flesch (1949), wrote: “[T]o most people, readability means ease of reading plus interest. They want to make as little effort as possible while they are reading, and they also want something ‘built in’ that will automatically carry them forward like an escalator” (p. 158). When the developers of another well-known formula, Edgar Dale and Jeanne S. Chall, “revisited” their earlier work in 1995, they provided their own somewhat expanded definition: “Readability is the sum total (including the interactions) of all those elements within a given piece of printed material that affect the success a group of readers have [*sic*] with it. The

success is the extent to which they understand it, read it at an optimal speed, and find it interesting” (Chall & Dale, 1995, p. 80).

Flesch (1948) and Dale and Chall (1948) developed readability formulas during the 1940s in part because scholars discovered that publishers, teachers, and individual readers could not assess reading difficulty levels accurately. Decades later, only five of 56 professional writers were able to rank five passages in the correct order of difficulty (Klare, 1976), a phenomenon corroborated by subsequent researchers (e.g., Hamilton & Shinn, 2003). In the 21st century, Sperling (2006) examined material that students had difficulty comprehending, material ostensibly at a grade 7 level. He used two different readability formulas to establish and verify that the material was actually written at a grade 9 level.

Outside of K-12 education, Friedman and Hoffman-Goetz (2006) reviewed 16 studies on the readability and comprehension levels of print and web-based materials written for cancer patients. They reported a range of mean reading difficulty levels from grade 6 to grade 14.1, whereas experts recommend readability for printed medical information at a grade 5 or grade 6 level. Weeks and Wallace (2002) concluded that articles in selected British medical journals were more readable than articles in selected American journals, but that both were difficult.² Finally, Harrington and Follett (1984) concluded that the readability levels of several child personality assessment instruments were too high for many young children.

Indeed, studies of materials from various fields have shown that reading difficulty levels are too high for many readers. This may be especially true in the United States, where many residents are not native English speakers. For example, Gunning (2000) and Maslin (2007) believe that difficult reading material contributes to negative attitudes toward reading among many children. Lewis, Colvard, and Adams (2008) calculated that 88% of examined privacy policy statements from banks were at a grade 12 or higher level of reading difficulty, while the average reading level of American high school graduates is grade 9; the average reading level for all American adults is only grade 7, according to DuBay (2004). It is probably not coincidental that the magazine with the world’s largest number of paid subscribers, *Reader’s Digest*, was written at a grade 8 or grade 9 level during its years of “great growth

[1930s-40s],” although by 1968 the difficulty level had increased to grade 10 (Gunning, 1968, p. 23).

Measures of readability have been applied to literature in other fields, and are used widely in education and (increasingly) by business and government (Fusaro, 1988). There have been several instances of courts of law accepting data derived from readability formulas as evidence, sometimes in tandem with testimony from readability experts. Worldwide use is confirmed by the existence of readability formulas for the Spanish, French, German, Dutch, Swedish, Russian, Hebrew, Hindi, Chinese, Vietnamese, and Korean languages (DuBay, 2004).

Despite their widespread use, there appear to be no published studies on the readability of materials in music education. Therefore, the primary purpose of the present study was to measure the readability levels of recent articles from two leading scholarly journals in the field. Secondary purposes were to compare readability results from five well-known readability formulas, and to compare the readability levels between the two journals and their four editors, two for each journal. We examined selected scholarly articles on the history of music education because: (a) the articles were available in electronic (Word) format; and (b) the articles were relatively free of mathematical symbols, tables, graphs, musical examples, photographs, lists, and other non-text content, all of which facilitated the analysis.

Method

Samples

The articles for this study were selected from the *JRME*, the world’s oldest scholarly journal in the field of music education, and the *JHRME*, the world’s only journal devoted exclusively to historical research in music education. The *JRME* has published articles based on various research methodologies since its founding in 1953. Yarbrough (2002) reported that through the journal’s first 50 years (1953-2002), 14.06% of the articles were historical ($n = 158$), which she defined as “research relating information from the past through the examination of documents and artifacts” (Yarbrough, 1984, p. 217). The *JHRME* (founded as the *BHRME* in 1980) has published only historical articles, plus editorials, book reviews, and a few miscellaneous pieces.

The present authors examined articles from the ten most recent calendar years (2000-09) prior to the period of data collection for this study. The *JRME* published 15 historical articles during that decade

² The statistical analysis reported in this article makes interpretation not only difficult, but questionable.

(Vol. 48, no. 1 to Vol. 57, no. 3).³ Four of those articles were based largely on quantitative methodology, which left 11 articles based on traditional historical methodology and reported in a narrative style. During that same decade, the *JHRME* published 88 historical articles (Vol. 21, no. 2 to Vol. 31, no. 1), one of which was based on quantitative methodology. We randomly selected 11 of the remaining 87 articles. These procedures resulted in two samples of equal size, one from each journal, for a total of 22 articles (see Appendix A for citations).⁴

Because editors of scholarly journals typically edit manuscripts at least to some extent, we examined the variable of editor in addition to the variable of journal. The first six articles in the *JRME* sample were published under the editorship of Cornelia Yarbrough, and the remaining five under editor Wendy Sims. The first six articles in the *JHRME* sample were published under the editorship of Jere Humphreys, and the remaining five under editor Mark Fonder. Most of the 22 articles were written by single authors ($n = 18$, 82%), and the others were written by two ($n = 2$, 9%), three ($n = 1$, 4.5%), or four ($n = 1$, 4.5%) authors. Some 21 different authors were represented in the two combined samples, five authors' names appeared on more than one article, and three authors were represented in both samples.

Formulas

Five readability formulas were utilized in this study: Flesch Reading Ease Score, Flesch-Kincaid Grade Level Index, Fog Index, SMOG Readability Formula, and New Dale-Chall Readability Formula.⁵ The oldest of the five, the Flesch Reading Ease Score, was a well-known formula developed by Rudolf Flesch in the 1940s. The scale ranged from 0-100, with higher numbers representing larger numbers of people who could comprehend the material. Scores from 0-29 represented very difficult material, 30-49 difficult, 50-59 fairly difficult, 60-69 standard, 70-79 fairly

easy, 80-89 easy, and 90-100 very easy. The Flesch-Kincaid Grade Level Index, developed in 1975, converted Flesch Reading Ease Scores into U.S. grade-level equivalents. Both Flesch formulas were based on the number of syllables in 100 words and on average sentence length.⁶

The third formula we applied was the Fog Index, published by Robert Gunning in 1952. It was based on three factors: sentence length, number of references to people, and number of prefixes and suffixes.⁷ The fourth formula, the SMOG Readability Formula, was related to the Fog Index. Developed by G. Harry McLaughlin in 1969, it was based on the number of words with three or more syllables in 30 sentences of text.⁸

The fifth formula employed, the New Dale-Chall Readability Formula, was a 1995 revision of a vocabulary-based readability formula developed by Edgar Dale and Jeanne Chall (1995) in 1948. This formula compared words in the text to a list of 3,000 "familiar" words, defined as words known to at least 80% of 4th-grade students. The formula also accounted for average sentence length. Grade-level readability equivalency was computed via a pair of

³ The volume and issue numbering scheme for the *JRME* changed; hence, 39 rather than 40 issues of this quarterly journal during the decade under investigation.

⁴ Technically, the 11 articles from the *JRME* did not constitute a sample, but instead were the entire corpus of articles from that decade that met the selection criteria for this study. It is referred to as a sample herein as a matter of convenience.

⁵ For general descriptions of selected readability formulas, see DuBay (2004), Burke and Greenberg (2010), and McLaughlin (1974).

⁶ The formula for the Flesch Reading Ease Score was: $206.835 - (1.015 \times \text{ASL}) - (84.6 \times \text{ASW})$, where ASL = number of words divided by number of sentences and ASW = number of syllables divided by number of words (Dalecki & Lasorsa, & Lewis, 2009). The equivalency formula for the Flesch-Kincaid Grade Level Index is: $\text{Grade level} = (.4 \text{ ASL}) + (12 \text{ ASW}) - 15$, where ASL = average sentence length and ASW = average number of syllables per word (DuBay, 2004).

⁷ The formula for the Fog Index (sometimes referred to as the "Gunning Fog Index") was: $\text{Grade level} = 0.4 (\text{average sentence length} + \text{percentage of Hard Words})$, where Hard Words = number of words with more than two syllables (Gunning, 1968). "Fog" refers to the "fog" that Robert Gunning perceived in newspaper writing (ca. 1952).

⁸ The SMOG Readability Formula was: $\text{Grade level} = 3 + \text{square root of total number of words with three or more syllables (in a sample of 30 sentences)}$. Like Gunning (1968), McLaughlin (1969) was concerned with newspaper (and other) writing. "SMOG" was his acronym for "Simple Measure of Gobbledygook" (p. 639).

conversion formulas.⁹ All five of these were at least two-variable formulas; in other words, they relied on computations of combinations of word and sentence variables.

As a measure of reliability, researchers have determined that these five and selected other readability formulas correlate positively with each other. For example, Pearson correlations of $r = .88 - .96$ were reported among the Flesch, Fog, SMOG, Dale-Chall, and one other formula when applied to “school-based literature,” and correlations of $r = .74 - .99$ were obtained for “health-based literature” (see Meade & Smith, 1991, p. 154). As for validity, studies showed that readability formulas predict a substantial (r^2) 50 – 84% of variance in reading comprehension (DuBay, 2004).

According to Shuptrine and Lichtenstein (1985), the New Dale-Chall Readability Formula measured the reading difficulty at which 50% of people in a given grade could read the material with understanding, whereas the two Flesch formulas were calibrated at 75% of people in a given grade reading with comprehension. The Fog and SMOG formulas measured comprehension differently from the Dale-Chall and Flesch formulas. The Fog Index required 90% comprehension of the text for an average individual at a given grade level, while the SMOG Readability Formula required 100% comprehension by an average person at a given grade level. Dubay (2004) explained that formulas based on relatively high criterion percentages (e.g., Fog and SMOG) tended to compute higher reading grade-level scores than did formulas based on lower criterion percentages (e.g., Dale-Chall and both Flesch formulas).

For various reasons we did not include any of the hundreds of other available readability formulas in this study. For example, the Spache (1953) formula was similar to the Dale-Chall formulas in its use of sentence length and percentage of difficult words, but it was designed to measure only up to grade 3 in difficulty. We also decided against using the FORCAST formula, which is designed for non-narrative documents such as

questionnaires, checklists, and job applications (Burke & Greenberg, 2010).

We selected three 100-word passages from each article in the combined samples ($N = 22$), and removed all markings that could have affected the calculation of sentence length and number of syllables in words, such as footnote markings and periods other than those at the ends of sentences (see Appendix B for the data used in the calculations). We then applied each formula, via software programs on a commercially available compact disk, to the selected passages (*Readability Formulas 7.4 and Dale-Chall Power 3.0*, 2011).

Results

Small sample sizes and lack of homogeneity of variances (Levine’s $F: p < .05$) prevented the use of parametric statistical tests. Therefore, non-parametric tests were employed, starting with the Kendall Coefficient of Concordance, which revealed a strong, statistically significant correlation among the four grade-level readability formulas ($W = .94, N = 22, df = 3, p < .01; p < .001$ before Bonferroni adjustment for the two Kendall runs). Adding the fifth formula, the Flesch Reading Ease Score, to the analysis model resulted in an even higher correlation coefficient ($W = .97, N = 22, df = 4, p < .01; p < .001$ before Bonferroni adjustment). A series of Spearman Coefficient of Rank Correlation computations, one for each pair among the five formulas, resulted in coefficients ranging from $\rho = .83$ to $.98$, all significant at the $p < .05$ level ($p < .01$ before Bonferroni adjustment for the ten Spearman runs) (see Table 1). Thus, not only did the five formulas correlate among themselves collectively, each individual formula correlated strongly with every other formula.

Table 1
Spearman Correlations (ρ ’s) between Formulas ($N = 22$)

	Flesch Ease	Flesch-Kincaid	Fog	SMOG	Dale-Chall
^a Flesch Ease	1.00	-.91*	-.84*	-.90*	-.83*
Flesch-Kincaid		1.00	.94*	.94*	.89*
Fog			1.00	.98*	.83*
SMOG				1.00	.85*
Dale-Chall					1.00

* $p < .05$ with Bonferroni adjustment for ten runs ($N = 22$) requires $\rho > .53, p < .005$.

^aCorrelations involving Flesch Reading Ease Scores are negative because the scale is reversed.

⁹ The original Dale-Chall Readability Formula was: Raw score = $0.1579 (PDW) + 0.0496 (ASL) + 3.6365$, where PDW = percentage of “Difficult Words” (i.e., not on the Dale-Chall list) and ASL = average sentence length. Raw scores are converted to grade levels via a chart (Dale & Chall, 1948). See Chall and Dale (1995) for the New Dale-Chall Readability Formula, conversion charts, and an updated list of “familiar” words. See also Fusaro (1988).

Results from the five readability formulas for each of the 22 articles are displayed in Table 2. Flesch Reading Ease results are on a 0-100 point scale, with lower numbers representing more difficult levels of readability. The mean of the Flesch Reading Ease Scores for all 22 articles was 38.7 (*SD* = 7.11),

which represents a relatively difficult readability level. The results for the other four formulas are given in equivalent U.S. school grades. Dale-Chall results are reported in grade ranges instead of single grade levels, so we provided the mean grade level of the reported range for each article.

Table 2
Results of Five Readability Formulas

Article	Formula						Grand Mean ^a
	Flesch Ease	Flesch-Kincaid	Fog	SMOG	Dale-Chall	D-C Mean	
JRME 1	24.0	18.5	21.6	17.9	16+	16	18.50
JRME 2	52.0	10.7	12.5	12.4	9-10	9.5	11.28
JRME 3	44.0	12.3	14.7	13.4	9-10	9.5	12.48
JRME 4	31.0	14.7	16.4	14.9	13-15	14.0	15.00
JRME 5	49.0	11.3	13.6	13.1	9-10	9.5	11.88
JRME 6	35.0	13.9	17.1	15.5	11-12	11.5	14.50
JRME 7	42.0	14.4	16.6	14.0	11-12	11.5	14.13
JRME 8	34.0	16.7	19.7	16.4	13-15	14.0	16.70
JRME 9	50.0	11.4	13.8	12.9	9-10	9.5	11.90
JRME 10	28.0	18.1	21.1	17.0	13-15	14.0	17.55
JRME 11	35.0	14.9	17.8	15.9	11-12	11.5	15.03
JRME Mean	38.5	14.3	16.8	14.9		11.9	14.45
JRME SD	9.4	2.7	3.1	1.8		2.3	2.43
JHRME 1	34.0	14.5	16.8	15.0	13-15	14.0	15.08
JHRME 2	45.0	12.3	15.2	13.5	9-10	9.5	12.63
JHRME 3	39.0	12.4	14.4	13.2	9-10	9.5	12.38
JHRME 4	32.0	14.5	17.5	15.4	11-12	11.5	14.73
JHRME 5	39.5	12.7	14.7	13.6	9-10	9.5	12.63
JHRME 6	38.0	12.4	15.3	14.2	11-12	11.5	13.35
JHRME 7	44.0	12.8	15.6	13.7	11-12	11.5	13.40
JHRME 8	42.0	13.4	16.6	14.7	11-12	11.5	14.05
JHRME 9	37.0	13.5	16.0	14.7	11-12	11.5	13.93
JHRME 10	42.0	12.6	14.7	13.5	11-12	11.5	13.08
JHRME 11	35.0	14.4	17.0	15.4	11-12	11.5	14.58
JHRME Mean	38.8	13.2	15.8	14.3		11.2	13.62
JHRME SD	4.2	0.9	1.1	0.8		1.3	0.92
Grand Mean	38.68	13.75	16.30	14.56		11.52	14.04
Grand SD	7.11	2.03	2.29	1.41		1.87	1.84
Mean Rank		2.18	4.00	2.28		1.00	
Range							11.28-18.50

^a Grand means based on the four grade-level formula results (excluding Flesch Ease), using means of the Dale-Chall grade-level ranges.

Although the four grade-level formulas correlated with each other, the resulting grade levels were significantly different from each other. As shown in Table 2, the highest mean for the 22 articles

came from the Fog Index (16.30), followed in descending order by the SMOG (14.56), Flesch-Kincaid (13.75), and Dale-Chall (11.52) formulas. A Friedman Two-Way Analysis of Variance By Ranks

Test revealed significant differences ($\chi^2 = 62.07, p < .001$) among the mean ranks (MR) produced by these four formulas: Fog (MR = 4.00), SMOG (MR = 2.82), Flesch-Kincaid (MR = 2.18), and Dale-Chall (MR = 1.00). The same order obtained for the two separate samples and for the combined samples.

The grand mean of the four grade-level formulas for all 22 articles was 14.04 years of schooling, near the beginning of the second year of college. There was a small standard deviation of only 1.84 (grade-level years). The mean of the four grade-level formulas for the most difficult article was 18.5, mid-way through the second year of graduate school. The lowest mean grade reading level was 11.28, approximately one quarter into the junior year of high school. Thus, despite the small standard deviation for the combined samples, the range of mean reading difficulty was quite large at 7.22 grade levels.

The mean readability level for the four grade-level formulas for the *JRME* sample ($M = 14.45$) was almost one grade level higher (.83) than the mean for the *JHRME* sample ($M = 13.62$) (see Table 2), a difference that failed to reach statistical significance (Mann Whitney U, $p > .05$). This non-significant difference was confirmed by five additional probability tests, one for each formula between the two journals (Mann Whitney U, $p > .05$). Similarly, a series of six probability tests revealed non-significant reading difficulty level differences for articles published under the direction of the four editors: five for the individual formulas and one for the five formulas combined (Kruskal-Wallis ANOVA, $p > .05$).

Conclusions

There has been some concern expressed in the literature over the validity of readability formulas; that is, the extent to which the formulas measure what they are purported to measure. Among other concerns, scholars (e.g., Bailin & Grafstein, 2001) have questioned the linguistic assumptions underlying readability formulas, while G. Harry McGlaughlin (1974), the developer of the SMOG Readability Formula, believes that they predict comprehension only. It should be noted that in the present study we did not “test the tests” *per se*, something scholars typically do in studies involving questions of test validity (see Anastasi & Urina, 1997). Instead, we attempted to measure a construct: the readability of samples of recent historical articles from two well-known music education research journals. In other words, we analyzed a body of written material, not a test instrument or human subjects.

Nevertheless, the fact that the readability formulas correlated strongly with each other in this study suggests that they were measuring the same construct. Furthermore, the fact that these formulas have been shown in other studies to correlate with reading comprehension suggests that the construct of readability, as measured by these formulas, is related to comprehension. On another level, limitations to the validity of the readability formulas might include the fact that the features of text that they measure, primarily lengths of sentences and words, may not reveal anything about the coherence or content of the text. However, except in purposefully extreme cases, such as a text composed of words in random order, readability experts have concluded that despite the raising of various challenges to the reliability and validity of readability formulas over the decades, “the preponderance of evidence generally supports their diagnostic value” (Dalecki, Lasorsa, & Lewis, 2009, p. 3). DuBay (2004) reported that as early as the 1980s there were some 200 readability formulas and more than 1,000 studies “attesting to their strong theoretical and statistical validity” (p. 2).

In the present study, the strong correlations among the five readability formulas ($W = .97$; $\rho = .83 - .98$) align with findings from studies of school-based ($r = .88 - .96$) and health-based ($r = .75 - .99$) reading materials (Meade & Smith, 1991). The fact that all five formulas were highly correlated in this study suggests that they may have measured similar traits, in this case the readability of selected journal articles on the history of music education.

Although the formula results were highly correlated, they computed different readability levels, which is also similar to findings from other fields, such as health care, where differences of up to 5.18 grade levels have been reported (Meade & Smith, 1991). An even wider range of mean grade levels was found in this study, based on the four grade-level formulas (grades 11.28 – 18.50) and confirmed by significant differences among the mean ranks of the formulas ($\chi^2 = 62.07, p < .001$). Further, the rank order of the formulas aligned with much of the research literature on the topic, with higher difficulty level results from the Fog and SMOG formulas, which are based on higher criterion values for knowledge or comprehension, and lower difficulty level results from the Flesch-Kincaid and Dale-Chall formulas, which are based on lower criterion values.

Regardless of grade-level differences among the formulas, the mean grade level resulting from the four grade-level formulas for all 22 articles examined in this study was 14.04. That suggests that the mean readability level was near the beginning of the second year of college. The mean difficulty levels ranged

from early in the junior year of high school (11.28) to mid-way through the second year of graduate school (18.50). Despite this wide range in readability, however, the small standard deviation of 1.84 means that approximately 68% of the articles in the combined samples (approximately 15 articles) fell within the range of early in the senior year of high school to late in the third year of college [($M = 14.04$) +/- ($SD = 1.84$) = 12.20 – 15.88 years of schooling]. That is, the standard deviation reveals that most of the articles fell within the range of late high school to near the end of the third year in college. The remainder fell outside that range, with one lying beyond a master's degree level of readability (18.50).

It appears that most of the articles examined in this study are at appropriate levels of readability for the presumptive readers of the two journals: graduate students, practicing teachers, and professors, all holders of bachelor's degrees and many with graduate degrees. However, because most people read at levels below their highest completed grade in school and also prefer to read easier material, many undergraduates could find some of the articles beyond their comfortable readability levels. On the other hand, this relatively difficult reading level may be necessary, because a small body of research suggests that relatively complex topics result in reports that are by their nature more difficult to read than most other types of works. One interesting study found that journalists who described actual, complex events wrote more complex articles at more difficult (i.e., lower) levels of readability than did journalists who had been discredited for their superficial, sometimes even fabricated reporting and writing (Dalecki, Lasorsa, & Lewis, 2009).

Nevertheless, many reading experts advocate the use of easy reading materials. Some government agencies now mandate easier reading levels for insurance, medical, tax, and other types of information aimed toward general readers. Moreover, it appears that reading experts have not advocated publicly for more difficult reading levels, in general or in specific media or subject areas. Similarly, there is no evidence that reading experts lament the lowering of reading difficulty levels (i.e., increasing of readability) of newspapers over the last several decades. On the contrary, reading experts do not view decreases in reading difficulty levels as a "dumbing down" of material to accommodate allegedly increasing numbers of poor readers. Instead, improving readability is seen as a positive trend

because research has shown that less difficult material can attract larger numbers of readers, and because people can read it with more comprehension. According to DuBay (2004), Robert Flesch and Robert Gunning, who worked with the Associated Press (AP) and United Press (now UPI), respectively, "had an enormous impact on journalism ... Together, they helped to bring down the reading grade level of front-page stories from the 16th to the 11th grade, where they remain today" (p. 23).

All five formulas employed in this study are relatively easy to compute, whether via software or manually, although preparing the materials for either type of analysis can be tedious. Because the Flesch Reading Ease and Flesch-Kincaid formulas are available on Microsoft Word programs, they could be the formulas of choice. Their use requires material in Word format, and extraneous characters should be removed. It should be kept in mind that these two formulas tend to score on the easier end of the readability range. Finally, Chall and Dale (1995) point out that "No readability formula is a complete and full measure of text difficulty" (p. 6).

The insights from this study and any future studies could help professors diagnose students' reading difficulties that may manifest themselves in inadequate reading comprehension and/or reluctance to read. Moreover, some sense of the reading level of materials could help professors, librarians, and others guide students toward materials at appropriate levels of readability, as well as help librarians responsible for collection development (Gray, 2012). In addition, the authors, reviewers, and editors responsible for producing the *JRME* and *JHRME* may gain some comfort from the knowledge learning that most of the selected articles seem to be at an appropriate level of readability for the intended readers. It should be kept in mind that any future increases in reading difficulty levels could result in smaller numbers of readers as well as lower levels of comprehension.

Future scholars could analyze the *JRME* and *JHRME* for changes in readability over time, as well as the readability of books, doctoral dissertations, and articles from other journals. Other materials related to music education could be examined also, including those aimed toward K-12 and undergraduate students. Finally, hypotheses about higher readability levels being related to increased readership and reading comprehension could be tested empirically with on a variety of music education reading materials.

REFERENCES

- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Bailin, A., & Grafstein, A. (2001). The linguistic assumptions underlying readability formulae: A critique. *Language & Communication, 21*(3), 285-301.
- Burke, V., & Greenberg, D. (2010). Determining readability: How to select and apply easy-to-use readability formulas to assess the difficulty of adult literacy materials. *Adult Basic Education and Literacy Journal, 4*(1), 34-42.
- Butler, D. M. (1973). An historical investigation and bibliography of nineteenth century music psychology literature, vols. I, II. Doctoral Dissertation, Ohio State University.
- Chall, J. S., & Dale, E. (1995). *Readability revisited: The New Dale-Chall Readability Formula*. Cambridge, Mass.: Brookline Books.
- Cooper, S., & Bayless, R. (2008). Examining the Music Teachers National Association papers and proceedings 1906 to 1930. *Journal of Historical Research in Music Education, 29*(2), 129-144.
- Dale, E., & Chall, J. S. (1948). A formula for predicting readability. *Educational Research Bulletin, 27*(2), 11-28.
- Dalecki, L., Lasorsa, D. L., & Lewis, S. C. (2009). The news readability problem. *Journalism Practice, 3*(1), 1-12.
- DuBay, W. H. (2004). *The principles of readability*. Impact Information: Costa Mesa, CA. Available at <http://www.impact-information.com/impactinfo/readability02.pdf>. ERIC (abstract): ED490073.
- Ebie, B. D. (2002). Characteristics of 50 years of research samples found in the *Journal of Research in Music Education, 1953-2002*. *Journal of Research in Music Education, 50*(4), 280-291.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*(3), 221-233.
- Flesch, R. (1949). *The art of readable writing*. New York: Harper & Brothers Publishers.
- Friedman, D. B., & Hoffman-Goetz, L. (2006). A systematic review of readability and comprehension instruments used for print and web-based cancer information. *Health Education & Behavior, 33*(3), 352-373.
- Fusaro, J. A. (1988). Applying statistical rigor to a validation study of the Fry Readability Graph. *Reading Research and Instruction, 28*(1), 44-48.
- Gray, C. J. (2012). Readability: A factor in student research? *The Reference Librarian, 53*(2), 194-205.
- Gunning, R. (1968). *The technique of clear writing* (Rev ed.). New York: McGraw-Hill Book Company.
- Gunning, T. G. (2000). *Best books for building literacy for elementary school children*. Boston: Allyn and Bacon.
- Hall, W. R. (1998). Twenty-five years of music research contributions. *Contributions to Music Education, 25*, 24-43.
- Hamilton, C., & Shinn, M. R. (2003). Characteristics of word callers: An investigation of the accuracy of teachers' judgments of reading comprehension and oral reading skills. *School Psychology Review, 32*(2), 228-240.
- Harrington, R. G., & Follett, G. M. (1984). The readability of child personality assessment instruments. *Journal of Psychoeducational Assessment, 2*(1), 37-48.
- Hedden, S.K. (1992). Research sessions at MENC conferences: 1970-1990. *The Quarterly Journal of Music Teaching and Learning, 3*(1), 80-85.
- Hedden, S. K. (1993). Music education research: A dozen conventions and a dozen JRME volumes. *The Bulletin of Historical Research in Music Education, 15*(1), 17-30.
- Humphreys, J. T. (1999). *The Bulletin of Historical Research in Music Education: Its authors, reviewers, and editorial committee members for the first twenty years*. *The Bulletin of Historical Research in Music Education, 20*(3), 171-180.
- Humphreys, J. T., Bess, D. M., & Bergee, M. J. (1997). Doctoral dissertations on the history of music education and music therapy. *The Quarterly Journal of Music Teaching and Learning, 7*(2-4), 114-124.
- Humphreys, J. T., & Stauffer, S. L. (2000). An analysis of the editorial committee of the *Journal of Research in Music Education, 1953-1992*. *Journal of Research in Music Education, 48*, 65-77.
- Kantorski, V. J. (1995). A content analysis of doctoral research in string education. *Journal of Research in Music Education, 43*(4), 288-297.
- Klare, G. R. (1976). Judging readability. *Instructional Science, 5*(1), 55-61.
- Kratus, J. (1992). Subjects in music education research. *The Quarterly Journal of Music Teaching and Learning, 3*(4), 50-54.
- Latimer, Jr., M. E. (2001). A history and analysis of the *Choral Journal* Editorial Board (1959-2009). *Journal of Historical Research in Music Education, 33*(1), 27-45.
- Lewis, S. D., Colvard, R. G., & Adams, C. N. (2008). A comparison of the readability of privacy statements of banks, credit counseling companies, and check cashing companies. *Journal of Organizational Culture, Communications and Conflict, 12*(2), 87-93.

- Lively, B. A., & Pressey, S. L. (1923). A method for measuring the 'vocabulary burden' of textbooks. *Educational Administration and Supervision*, 9, 389-398.
- Maslin, P. (2007). Comparison of readability and decodability levels across five first grade basal programs. *Reading Improvement*, 44(2), 59-75.
- McCarthy, M. (1999). *The Bulletin of Historical Research in Music Education: A content analysis of articles in the first twenty volumes. The Bulletin of Historical Research in Music Education*, 20(3), 181-202.
- McLaughlin, G. H. (1969). SMOG Grading--a new readability formula. *Journal of Reading*, 12(8), 639-646.
- McLaughlin, G. H. (1974). Temptations of the Flesch. *Instructional Science*, 2(4), 367-383.
- Meade, C. D., & Smith, C. F. (1991). Readability formulas: Cautions and criteria. *Patient Education and Counseling*, 17(2), 153-158.
- Preston, K. Y., & Humphreys, J. T. (2007). Historical research on music education and music therapy: Doctoral dissertations of the twentieth century. *Journal of Historical Research in Music Education*, 29(1), 55-73.
- Readability Formulas 7.4 and Dale-Chall POWER 3.0* (2011, copyright). Micro Power & Light, compact disk. Available at <http://www.micropowerandlight.com/rd.html>.
- Shuptrine, F. K., & Lichtenstein, D. R. (1985). Measuring readability levels of undergraduate marketing textbooks. *Journal of Marketing Education*, 7(3), 38-45.
- Spache, G. (1953). A new readability formula for primary-grade reading materials. *The Elementary School Journal*, 53(7), 410-413.
- Sperling, R. A. (2006). Assessing reading materials for students who are learning disabled. *Intervention in School and Clinic*, 41(3), 138-143.
- Stabler, D. G. (1986). A content analysis of the *Bulletin of the Council for Research in Music Education*, 1963-1985. Doctoral Dissertation, University of Illinois at Urbana-Champaign.
- Weeks, W. B., & Wallace, A. E. (2002). Readability of British and American medical prose at the start of the 21st century. *British Medical Journal*, 325, 1451-1452.
- Yarbrough, C. (1984). A content analysis of the *Journal of Research in Music Education*, 1953-1983. *Journal of Research in Music Education*, 32(4), 213-222.
- Yarbrough, C. (2002). The first 50 years of the *Journal of Research in Music Education: A content analysis. Journal of Research in Music Education*, 50(4), 276-279.

CHINESE ABSTRACT

中文摘要

《音樂教育研究期刊》與《音樂教育歷史研究期刊》中部分文章的閱讀難度水準

Alexandra H. Humphreys

Arizona State University—Phoenix, U.S.A.

Jere T. Humphreys

Arizona State University—Tempe, U.S.A.

可讀性估公式被廣泛運用在教育領域，而且在商業和政府部門的使用頻率與日俱增。在過去的 30 年裏，研究人員共研發了 200 多種可讀性估公式，這些公式所估算的可讀性指標與實際閱讀理解程度具有中度或很強的相關性。本文使用了五種著名的可讀性估公式來評估最近發表的 22 篇音樂教育史文章，發現這五種公式的計算結果有很強的相關性。這 22 篇文章的平均閱讀難度水準是 14.04，接近大學二年級初期的閱讀水準。由於先前的研究結果顯示大部分人喜歡閱讀難度水準低於本人學歷水準的文章以及比較容易理解的文章，這兩種期刊中文章的閱讀難度水準非常適合預計的閱讀對象（有學士學位和研究生學位的人）。本文研究結果可能會對那些大學教授，圖書館員和其他負責指導學生閱讀的人員有所幫助。

Appendix A

Articles Analyzed (2000-09) (N = 22)

Journal of Research in Music Education (JRME) (n = 11 articles, listed chronologically)

- 1 Campbell, P. S. (2000). How musical are we: John Blacking on music, education, and cultural understanding. *JRME*, 48, 336-359.
- 2 Volk, T. M. (2001). Little Red Songbooks: Songs for the labor force of America. *JRME*, 49, 33-48.
- 3 Howe, S. W. (2003). The NBC Music Appreciation Hour: Radio broadcasts of Walter Damrosch, 1928-1942. *JRME*, 51, 64-77.
- 4 Colwell, C. M., & Heller, G. N. (2003). Lowell Mason's *The Song Garden* (1864-66): Its background, content, and comparison to a twentieth-century series. *JRME*, 51, 231-244.
- 5 Howe, S. W. (2004). Elsie Shawe, music supervisor in St. Paul, Minnesota (1898-1933). *JRME*, 52, 328-342.
- 6 Cooper, S. (2005). Marguerite V. Hood and music education radio broadcasts in rural Montana (1937-39). *JRME*, 53, 295-307.
- 7 Hedden, D. G., Heller, G. N., Humphreys, J. T., & Slattery, V. A. (2007). Alice Carey Inskip (1875-1942): A pioneering Iowa music educator and MENC founding member." *JRME*, 55, 129-147.
- 8 Hash, P. M. (2007). The Chicago Reform School Band: 1862-1872. *JRME*, 55, 252-267.
- 9 Volk, T. M. (2007). "Charts and other paraphernalia": Charles H. Congdon (1856-1928) and his music teaching materials. *JRME*, 55, 302-312.
- 10 Watts, S. H., & Campbell, P. S. (2008). American folk songs for children: Ruth Crawford Seeger's contributions to music education. *JRME*, 56, 238-254.
- 11 Hash, P. M. (2009). The National High School Orchestra 1926-1938. *JRME*, 57, 50-72.

Journal of Historical Research in Music Education (JHRME) (n = 11 articles, listed chronologically)

- 1 Kennedy, M. A. (2000). Creative music making since the time of the singing schools: Fringe benefits. *JHRME*, 21, 132-148.
- 2 Heller, G. N. (2001). Allen Perdue Britton and "The Study of Music: An Academic Discipline." *JHRME*, 22, 94-109.
- 3 Kou, M-L. L. (2001). Development of music education in Taiwan (1895-1995). *JHRME*, 22, 177-190.
- 4 Lee, A. H-C. (2002). The influence of Japanese music education in Taiwan during the Japanese Protectorate. *JHRME*, 23, 106-118.
- 5 Fickett, J. H. (2002). A history of music education at Michigan State University. *JHRME*, 23, 119-136.
- 6 Aguilar, B., Ramsey, D., & Lumsden, B. (2002). The Aztec Empire and the Spanish missions: Early music education in North America. *JHRME*, 24, 62-82.
- 7 Colwell, R. (2006). Susanne Langer, Charles Leonhard, and augury. *JHRME*, 27, 132-144.
- 8 Russell, J. A. "The origin and use of an authentic Irish folk tune in American school orchestra arrangements. *JHRME*, 28, 38-52.
- 9 Lee, W. R. (2007). A new look at a significant cultural moment: The Music Supervisors National Conference 1907-1932. *JHRME*, 28, 93-110.
- 10 Volk, T. M. (2007). Anne Shaw Faulkner Oberndorfer (1877-1948): Music educator for the homemakers of America. *JHRME*, 29, 26-38.
- 11 Hash, P. M. (2008). History of the Illinois School Band Association: 1924-1941. *JHRME*, 30, 4- 20.

Appendix B

Data Used to Calculate Readability Levels of the Combined Samples of Articles ($N = 22$)

		Data Type						
Article		Words #	Syllables #	Monosyllable Words #	Monosyllable Words %	# Words >2 syl	%Words <2 syl	Difficult Words #
JRME 1	1	1,043	1,817	594	56.95	221	21.19	201
JRME 2	2	564	907	353	62.59	88	15.60	70
JRME 3	3	622	1,044	356	57.23	109	17.52	99
JRME 4	4	705	1,264	378	53.62	141	20.00	123
JRME 5	5	590	963	365	61.86	101	17.12	85
JRME 6	6	577	1,019	322	55.81	135	23.40	119
JRME 7	7	849	1,365	504	59.36	120	14.13	112
JRME 8	8	989	1,627	603	60.97	179	18.10	160
JRME 9	9	615	992	370	60.16	97	15.77	86
JRME 10	10	1,056	1,788	636	60.23	197	18.66	186
JRME 11	11	815	1,395	481	59.02	173	21.23	148

JHRME 1	1	717	1,261	404	56.35	145	20.22	130
JHRME 2	2	642	1,061	393	61.21	110	17.13	106
JHRME 3	3	554	977	281	50.72	103	18.59	97
JHRME 4	4	655	1,166	358	54.66	144	21.98	133
JHRME 5	5	592	1,031	327	55.24	112	18.92	100
JHRME 6	6	547	969	313	57.22	126	23.03	110
JHRME 7	7	699	1,146	434	62.09	114	16.31	109
JHRME 8	8	728	1,204	439	60.30	136	18.68	126
JHRME 9	9	648	1,132	380	58.64	137	21.14	119
JHRME 10	10	627	1,066	356	56.78	111	17.70	100
JHRME 11	11	715	1,252	412	57.62	154	21.54	134