

THE JOURNAL OF APPELLATE PRACTICE AND PROCESS

PRACTICE NOTE

DOES THE READABILITY OF YOUR BRIEF AFFECT
YOUR CHANCE OF WINNING AN APPEAL?

Lance N. Long and William F. Christensen*

I. INTRODUCTION

The short answer is “no”—at least if by “readability” you mean readability as judged by two of the several well-recognized readability formulas developed by researchers during the past fifty or sixty years.¹ Using the Flesch Reading Ease scale and the Flesch-Kincaid Grade-Level scale, we analyzed the readability of 882 state, federal, and United States Supreme Court briefs and found no statistically significant relationship

* Lance N. Long is an Associate Professor of Legal Skills at Stetson University College of Law and a Visiting Professor of Legal Research and Writing at the University of Oregon. William F. Christensen is a Professor of Statistics at Brigham Young University. We are thankful for the tireless efforts of our research assistants, Abigail Pressler, Aydin Bonabi, and Dallan Bunce, who compiled the data for our analysis; Alexander Zaitzeff, who assisted with some of the data analysis; and Emily Follansbee, who helped with some last-minute research. We are also thankful to Kirsten Davis, Tom Lee, John Fee, and Catherine Cameron for their ideas while this article was in its preliminary stages, and to Laura Beaton for her help in reading and editing drafts and footnotes. We acknowledge and thank Stetson University College of Law, Brigham Young University, and the University of Oregon for their support of this article.

1. See discussion in Part II, *infra*.

between the readability of those briefs and success on appeal. This is good news for those who would like to believe that appeals are decided on the merits of a case and that the success of an appeal is not influenced by the “readability” of a brief (or by any other writing convention, for that matter). It is also good news for critics of readability formulas, who may believe either that such formulas are fundamentally flawed² or that such formulas have little to do with legal writing.³ It provides readability critics with yet another argument for disregarding readability analyses (even if readability is valid, it doesn’t make a quantifiable difference). On the other hand, this may be less welcome news for legal writing professionals who may want to believe that the likelihood of success on appeal can be increased by writing a more “readable” brief and that a computerized readability formula can provide a basis for determining readability.⁴

Our study was an attempt to determine whether using long sentences and long words correlates with success on appeal. The Flesch readability formulas that we used in our study measure precisely those two elements. We were not particularly concerned about whether readability formulas accurately measure the appropriateness of a particular text for a particular reader. Appellate briefs, no matter how readable or unreadable they may be, are read by a highly educated audience. But for all readers, including highly educated readers, it is generally easier to read shorter sentences and shorter words.⁵ Could this possibly make a difference in the outcome of an appeal?

2. See n. 28, *infra*.

3. See n. 39, *infra*.

4. Of course, not all legal writing professionals share this belief. In particular, Professor Sirico has noted that unless readability formulas are properly understood and implemented, reliance on them can be counterproductive. See generally Louis J. Sirico, Jr., *Readability Studies: How Technocentrism Can Compromise Research and Legal Determinations*, 26 *Quinnipiac L. Rev.* 147 (2007). Our findings, although based on a formula that he criticizes, suggest that Flesch readability has little applicability to appellate brief success.

5. See generally Peter M. Tiersma, *Legal Language* 219–20 (U. of Chicago Press 1999) (noting that even appellate judges may be befuddled by the language of the typical insurance contract); Rudolf Flesch, *How to Write Plain English, A Book for Lawyers and Consumers* 20–22 (Harper & Row 1979) (discussing this phenomenon and concluding by noting that “the longer a sentence, the harder it is to read”); see also n. 38, *infra*.

Our study suggests that the length of sentences and words, which is “readability” for our purposes, probably does not make much difference in appellate brief writing. First, we found that most briefs are written at about the same level of readability; there simply is not much difference in how lawyers write appellate briefs when it comes to the length of sentences and words. Furthermore, the readability of most appellate briefs is well within the reading ability of the highly educated audience of appellate judges and justices. Second, the relatively small differences in readability are not related to the outcome of an appeal in a statistically significant manner. Our study did show, however, that the opinions of judges and justices are less readable than lawyers’ briefs and that the opinions of dissenting judges or justices are the least readable of all the appellate writing we analyzed. Ultimately, we conclude that readability, as determined by the Flesch Reading Ease scale, is a non-issue for legal writing at the appellate level.

The analysis discussed in this Article uses a methodology and approach similar to that used in the authors’ previously published article,⁶ which described “an empirical study of 800 federal and state appellate briefs randomly selected for the purpose of determining whether any relationship exists between intensifier use in the parties’ briefs and the outcomes in those cases.”⁷ This article utilizes a random selection of those same 800 briefs and adds to them the petitioner and respondent briefs written to the United States Supreme Court in every case in which the Court issued an opinion from the time Justice Alito joined the Court until shortly before the retirement of Justice Souter.

A readability study using the Flesch Reading Ease scale and the Flesch-Kincaid Grade scale was performed on the Supreme Court briefs and the randomly selected state and federal briefs, as well as the court opinions associated with those briefs. This Article examines the results of that analysis. Although readability did not appear to be related to outcome, there was a statistically significant relationship between the

6. Lance N. Long & William F. Christensen, *Clearly, Using Intensifiers Is Very Bad—Or Is It?* 45 Idaho L. Rev. 171 (2008).

7. *Id.* at 173. Also, note that, unlike the present study, the first study found a statistically significant relationship between the use of intensifiers and success on appeal.

readability of the courts' majority and dissenting opinions. Dissenting opinions are decidedly less readable than majority opinions.⁸

Part II of this Article discusses readability formulas generally and the Flesch Reading Ease scale and the Flesch-Kincaid Grade scale specifically, including the criticisms and limitations of readability formulas. Part III explains our analysis methodology and the results of our analysis. Part IV contains a discussion of our results and our conclusions.

II. READABILITY, READABILITY FORMULAS, AND LEGAL WRITING

It is beyond the scope of this article to fully examine and explain readability formulas or the history and theory behind those formulas.⁹ A short summary of the concept of readability, however, is provided here to explain the purpose of readability formulas and to explain why we applied one such formula to appellate briefs and opinions.

“Readability” is what makes some texts easier to read than others.”¹⁰ Since the 1920s, researchers, including linguists, educators, psychologists, and other scholars, have analyzed

8. A forthcoming article by the authors, tentatively titled *Why Losing Lawyers and Dissenting Judges Write Differently from the Winners—Intensifiers, Readability, and the Theory of Argumentative Threat*, will more fully discuss the relationship between the use of intensifiers and readability by winning and losing brief writers and majority and dissenting judges and justices.

9. At least one excellent and easily accessible short history and explanation of readability formulas and the theory underlying them can be found on the Web. See William H. DuBay, *The Principles of Readability*, <http://www.impact-information.com/impactinfo/readability02.pdf> (Aug. 25, 2004) (accessed Aug. 4, 2011; copy on file with Journal of Appellate Practice and Process). This paper includes an extensive bibliography of important and seminal works on readability. See *id.* at 59.

10. *Id.* at 3. DuBay also provides three additional definitions of “readability.” *Id.* A less readable, but perhaps more precise, definition of readability is offered by the creator of the SMOG (Simple Measure of Gobbledegook) readability formula, G. Harry McLaughlin:

One of the least ambiguous published definitions of readability is that given by English and English in their *Dictionary of Psychological Terms*. “Readability,” they say, “is the quality of a written or printed communication that makes it easy for any given class of persons to understand its meaning, or that induces them to continue reading.”

G. Harry McLaughlin, *Proposals for British Readability Measures*, in *The Third International Reading Symposium: Today's Child and Learning to Read* 186, 186 (John Downing & Amy L. Brown eds., Cassell 1968).

writing to determine what makes it more or less readable.¹¹ By the 1950s, several formulas for assessing readability had been developed,¹² and “[b]y the 1980s, there were 200 formulas and over a thousand studies published on the readability formulas attesting to their strong theoretical and statistical validity.”¹³

Each of the hundreds of readability formulas uses a different set of semantic and syntactic factors to determine readability, but the most frequently used factors are word complexity and sentence length.¹⁴ And while these rather simple “surface features” exclude any consideration of content, grammar, or organization, over fifty years of research have shown that these factors are the best predictors of readability based on comprehension tests that do consider content, grammar, and organization.¹⁵ Although almost every conceivable linguistic factor has been included in the scores of different formulas, and some formulas include a dozen or more factors, the addition of more factors does little to increase the accuracy of readability predictions and renders the formulas much more difficult to use.¹⁶ “Put another way, counting more things does not make [a] formula any more predictive of reading ease but takes a lot more effort.”¹⁷

11. See e.g. DuBay, *supra* n. 9, at 2–3; Cheryl Stephens, *All About Readability*, <http://www.plainlanguage.org/stephens/readability.html> (Plain Lang. Assn. Intl. 2000) (accessed Aug. 5, 2011; copy on file with Journal of Appellate Practice and Process); see also Jeanne S. Chall, *The Beginning Years*, in *Readability: Its Past, Present, and Future* 2, 2–4 (Beverly L. Zakulak & S. Jay Samuels eds., Intl. Reading Assn. 1988) (summarizing history of “readability measurement”); George R. Klare, *The Formative Years*, in *Readability: Its Past, Present, and Future*, *supra* this note, at 14, 14–23.

12. DuBay, *supra* n. 9, at 3, 13–31.

13. *Id.* at 2; see also Joseph Kimble, *Answering the Critics of Plain Language*, 5 *Scribes J. Leg. Writing* 51 (citing numerous studies showing that plain language improves comprehension).

14. See DuBay at 13–31 (listing studies); Stephens, *supra* n. 11.

15. *Id.* at 35–36; see also Jeanne S. Chall & Edgar Dale, *Readability Revisited: The New Dale-Chall Readability Formula 5–6* (Brookline Books 1995) (“[T]he strongest predictor of overall text difficulty [is] word difficulty The next best predictor of comprehension difficulty . . . is sentence length. Sentence length stands up quite well as a predictor of syntactic complexity—even better than more complex syntactic measures based on sophisticated linguistic theories.”) As noted *infra*, however, these claims have been questioned and criticized.

16. DuBay, *supra* n. 9, at 18–19.

17. Stephens, *supra* n. 11.

Of the many readability formulas, some of the more popular and accurate formulas that rely on sentence and word length include the SMOG (Simple Measure of Gobbledegook) formula, which “measures the number of words of more than 2 syllables in a sample of 30 words,”¹⁸ the Gunning Fog Index, which uses two variables: “in a sample of 100 words, the average number of words per sentence and the number of words of more than 2 syllables,”¹⁹ the Flesch Reading Ease formula, which measures “the number of syllables and the number of sentences for each 100-word sample,” and the Flesch-Kincaid Grade Level formula, which measures the same variables, but converts them to a grade level calculation.²⁰ Flesch Reading Ease scores range from 0 to 100; a score of 0 is practically unreadable, a score of 30 means the reading is “very difficult,” a score of 70 means the reading is suitable for adult audiences, and a score of 100 means the reading is easy²¹ and should be readable by someone with a fourth grade education who is “barely ‘functionally literate.’”²²

These formulas are popular because they are relatively easy to use (all four can be applied with readily available software),²³ and they appear to be accurate because they correlate well with more sophisticated, content-based measures of reading comprehension.²⁴ The Flesch Reading Ease formula is probably

18. Geoffrey Marnell, *Measuring Readability, Part 1: The Spirit is Willing, but the Flesch is Weak* 3, <http://www.abelard.com.au/readability%20statistics.pdf> (accessed Aug. 15, 2011; copy on file with Journal of Appellate Practice and Process); see also DuBay, *supra* n. 9, at 47.

19. Marnell, *supra* n. 18, at 3; see also DuBay, *supra* n. 9, at 24.

20. DuBay, *supra* n. 9, at 21, 50.

21. *Id.* at 22 (including chart).

22. *Id.* at 21 (quoting language used by U.S. Census in 1940s); see also Marnell, *supra* n. 18, at 3 (noting that a company writing materials designed to accompany a product to be imported to the United States “was asked to ensure that the documentation had a readability score that indicated that it could be fully understood by someone with only an eighth-grade education”). Of course, the Flesch Grade Level scale approximates the reading ability of a person in a given school grade. See e.g. DuBay, *supra* n. 9, at 50.

23. Thomas Oakland & Holly B. Lane, *Language, Reading, and Readability Formulas: Implications for Developing and Adapting Tests*, 4 *Intl. J. Testing* 239, 250 (2004).

24. DuBay, *supra* n. 9, at 22–24 (citing research), 36 (noting that “the readability variables . . . with all their limitations have remained the best predictors of text difficulty as measured by comprehension tests”); McLaughlin, *supra* n. 10, at 191–92 (concluding that the readability measures are sufficiently predictive to be useful). One such measure of reading comprehension is the Cloze test, developed by Wilson Taylor in 1953. The Cloze

the most influential and popular readability formula, due in part to its adoption by Microsoft Word.²⁵ The Flesch Reading Ease formula uses the following formula to determine readability: “Reading ease (RE) = $206.835 - 84.6s - 1.015w$ where s = the average number of syllables per word and w = the average number of words per sentence.”²⁶ The Flesch-Kincaid Grade Level formula uses the following formula to determine readability: Grade level (GL) = $.39s + 11.8w - 15.59$.²⁷ Therefore, our analysis considers only the relationship of word length and sentence length with success on appeal.

Even though readability formulas correlate with reading comprehension, this correlation has been questioned, and recently, the reliability of readability formulas has been criticized.²⁸ Some scholars and experts claim that readability

test deletes words from a text “(usually every fifth word) and requires the subjects to fill in the blanks.” A score is derived from the percentage of words correctly filled in by the subject. The Cloze test is the subject of more than a thousand studies, and is well accepted as a measure of reading comprehension. DuBay, *supra* n. 9, at 27–28; *but see* Stephens, *supra* n. 11 (citing studies indicating that Cloze testing is “more suitable to assess readers’ abilities than to measure the readability of text”).

25. Marnell, *supra* n. 18, at 1, 3; Sirico, *supra* n. 4, at 148.

26. Flesch, *supra* n. 5 at 24; *see also* Marnell, *supra* n. 18, at 3; DuBay, *supra* n. 9, at 21–22 (including illustrative chart).

27. DuBay, *supra* n. 9, at 50 (showing the formula as “GL = (.4ASL) + (12ASW) – 15”); Sirico, *supra* n. 4, at 151 (using the formula—both in Word and by hand—to compute score for language from a lease).

28. *See e.g.* Bertram C. Bruce, Ann D. Rubin, & Kathleen S. Starr, *Why Readability Formulas Fail* (U. of Ill. Ctr. for the Study of Reading 1981), <https://www.ideals.illinois.edu/bitstream/handle/2142/15490/why-rf-fail.pdf?sequence=2> (criticizing readability formulas for insufficient consideration of relevant factors, lack of statistical rigor, and inappropriate use) (accessed Aug. 15, 2011; copy on file with Journal of Appellate Practice and Process); Oakland & Lane, *supra* n. 23 (criticizing readability formulas when used by persons without expertise in reading and language); Emily Pitler & Ani Nenkova, *Revisiting Readability: A Unified Framework for Predicting Text Quality*, <http://www.cis.upenn.edu/~nenkova/papers/revisitingReadability.pdf> (accessed Aug. 15, 2011; copy on file with Journal of Appellate Practice and Process) (arguing that surface measures do not predict readability and proposing a sophisticated formula using the number of verb phrases, the length of the article, the likelihood of the vocabulary (suggesting the audience’s baseline familiarity with the words used), and the likelihood of discourse relations as a statistically valid alternative); Marnell, *supra* n. 18, at 4–8 (criticizing readability formulas, in particular the Flesch Reading Ease formula, as being flawed and failing to accurately predict readability); Norman Otto Stockmeyer, *Using Microsoft Word’s Readability Program*, 88 Mich. Bar J. 46 (Jan. 2009) (noting problems with Microsoft’s program and advising against exclusive reliance on the Flesch Reading Ease formula); Sirico, *supra* n. 4, at 151–52, 165–69 (questioning the validity of

depends more on the literacy, motivation, and background of the reader than the surface factors of the text.²⁹ Some further claim that reliance on such formulas can actually decrease the readability of text, especially when “writers . . . write to the formulas.”³⁰ By trying to lower reading difficulty through the use of shorter sentences, a writer can actually reduce the semantic flow of an idea and make it more difficult to understand.³¹

Furthermore, predicted readability of the same text can, and usually does, vary greatly between various formulas,³² and computerized versions of a given formula may not always faithfully execute the correct formula.³³ A particularly insightful criticism of computerized readability formulas in this regard was raised by Professor Sirico,³⁴ who claims that the Flesch formulas used by Microsoft Word (which are the formulas we used in our research) do not actually use the Flesch formulation at all, but seem to rely instead on “some algorithm to approximate the number of syllables.”³⁵ This is why there may be discrepancies between various versions of the Flesch formulations, as well as differences between computer calculations and hand calculations of the same formula.³⁶

Nevertheless, at a minimum, readability formulas can be a helpful tool for roughly gauging the difficulty of longer texts and for providing a measure for determining whether that difficulty has been reduced in the revision process.³⁷

Microsoft’s version of the Flesch Reading Ease formula, and explaining the perils of using the Flesch formula beyond its intended parameters).

29. See, e.g., DuBay, *supra* n. 9, at 28–31 (citing research); Bruce, *supra* n. 28, at 1–2; Oakland & Lane, *supra* n. 23, at 245–50.

30. Bruce et al., *supra* n. 28, at 3 (pointing out that “[s]uch prescriptive use magnifies the inaccuracies inherent in the formulas”); see also e.g. Oakland & Lane, *supra* n. 23, at 245–50; Klare, *supra* n. 11, at 27.

31. See e.g. Pitler & Nenkova, *supra* n. 28, at 5; Mamell, *supra* n. 18, at 4.

32. See DuBay, *supra* n. 9, at 55–56, for a discussion of this problem.

33. Sirico, *supra* n. 4, at 151–52; Stockmeyer, *supra* n. 28, at 47.

34. See generally Sirico, *supra* n. 4.

35. *Id.* at 165.

36. *Id.* at 165–66.

37. See Tiersma, *supra* n. 5, at 225–27; Stephens, *supra* n. 11; Stockmeyer, *supra* n. 28, at 47; see also Mary Ann Hogan, *Flesch and the Common Man: Why Foundation Bigs Should Use Little Words*, <http://www.knightcommunications.org/promotion-101/news-release-workshop/flesch-and-the-common-man/> (applying Flesch test to several news

Readability formulas in legal writing have been primarily applied to statutory and contract language, ballot measures, and jury instructions, usually in an effort to comply with legislation and administrative rules requiring the use of plain language, and studies in these areas generally show that more readable language is better understood than less readable language.³⁸ Until recently, however, little attention has been given to the readability of legal memoranda. The common wisdom seemed to hold that readability was not applicable to the sophisticated and complex nature of legal writing.³⁹ While studies have shown that plain English is preferred over legalese in legal memoranda,⁴⁰ until recently, there were no studies addressing Flesch-type readability and appellate briefs. Two recent studies that discussed readability in connection with the “Questions

releases) (accessed Aug. 15, 2011; copy on file with Journal of Appellate Practice and Process).

38. See e.g. Robert W. Benson & Joan B. Kessler, *Legalese v. Plain English: An Empirical Study of Persuasion and Credibility in Appellate Brief Writing*, 20 Loy. L.A. L. Rev. 301, 302 (1987); see generally e.g. Edward Fry, *The Legal Aspects of Readability*, <http://www.eric.ed.gov/PDFS/ED416466.pdf> (revised version of a speech given at the International Reading Association meeting in May 1989); Tiersma, *supra* n. 5, at 220–27; DuBay, *supra* n. 9, at 54–55; Robert W. Benson, *The End of Legalese: The Game is Over*, 13 N.Y. U. Rev. L. & Soc. Change 519, 547–58 (1984–85); see also Kimble, *supra* n. 13, at 68–71 (describing a study showing that contract and statutory provisions were better understood by law students, law school staff, and state-agency staff when the provisions were rewritten in a more readable format).

39. See James Lindgren, *Style Matters*, 92 Yale L.J. 161, 169 (1982) (characterizing Flesch’s then-new *How to Write Plain English*, *supra* n. 5, as “good,” but questioning the value of applying a Flesch-type analysis to legal writing and asking rhetorically: “Why force yourself to write at an eighth- or ninth-grade level if you are writing mainly for an audience of other lawyers?”). An example of this sentiment was also stated by the court in *Johnson v. Revenue Mgt. Corp.*, 169 F.3d 1057 (7th Cir. 1999), when scrutinizing dunning letters sent to debtors by collection agencies: “Unsophisticated readers may require more explanation than do federal judges; what seems pellucid to a judge, a legally sophisticated reader, may be opaque to someone whose formal education ended after the sixth grade.” *Id.* at 1060.

40. See e.g. Sean Flammer, *An Empirical Analysis of Writing Style, Persuasion, and the Use of Plain Language*, 16 Leg. Writing 183 (2010) (describing a survey showing that most state and federal judges prefer plain language over legalese and describing three earlier surveys that reached the same result); Hunter M. Breland & Frederick M. Hart, *Defining Legal Writing: An Empirical Analysis of the Legal Memorandum*, L. School Admission Council Research Rep. 93-06 (April 1994) (describing an extensive survey and regression analysis conducted to determine what constitutes good or poor legal writing); Benson & Kessler, *supra* n. 38; cf. Tiersma, *supra* n. 5, at 211–30 (listing areas in which plain language is better understood than unduly technical language and discussing examples).

Presented” in appellate briefs⁴¹ and a study of United States Supreme Court briefs and Flesch readability⁴² found no correlation between Flesch readability scores of the parties’ briefs and the outcome of appeal. However, to our knowledge, no other study has yet applied a regression analysis to determine whether readability of an appellate brief is related to the outcome of the appeal. Our study confirms the findings of Professor Coleman and Mr. Phung and shows no relationship between the readability of a brief and the outcome of an appeal.

For our purposes, the limitations and criticisms of readability formulas are largely irrelevant. We chose the Flesch formulas because we wanted to see if using longer sentences and longer words correlated with success on appeal. We assumed that the audience for appellate briefs could read the longer, more complex words and sentences. We only wanted to know whether the length of words and sentences correlated with success on appeal. For this limited purpose, the Flesch formulas are appropriate. Although we did not expect to find any relationship, we had previously found such a relationship in connection with the frequent use of intensifiers, and so, we gave it a go. As explained below, our hypothesis was confirmed, and we did not find any such relationship; however, we found some interesting trends and patterns that may be helpful to legal writers.

III. METHODOLOGY AND THE RESULTS OF THE ANALYSIS

A. The Study and Case Database

Notwithstanding the possible shortcomings of Microsoft Word’s version of the Flesch and Flesch-Kincaid formulations, because we did not need to compare our results with another formula and because we were using a formula only to obtain a consistent comparison of the briefs and opinions in the study, we believed that Microsoft Word’s version would be a good choice.

41. Judith D. Fischer, *Got Issues? An Empirical Study about Framing Them*, 6 J. Assoc. Leg. Writing Dirs. 1 (2009); Brady S. Coleman et al., *Grammatical and Structural Choices in Issue Framing: A Quantitative Analysis of “Questions Presented” from a Half Century of Supreme Court Briefs*, 29 Am. J. Tr. Advoc. 327 (2005).

42. Brady Coleman & Quy Phung, *The Language of Supreme Court Briefs: A Large-Scale Quantitative Investigation*, 11 J. App. Prac. & Process 75 (2010).

Furthermore, the ease of use and the accessibility of Microsoft's software to anyone wanting to compare our results with their own writing also made the Word version a good choice. We used the Flesch Reading Ease formula and the Flesch-Kincaid formula from Word 2007 to ensure that we would get the full possible range of results in connection with the Flesch-Grade scale.⁴³

Our study database consisted of court opinions and briefs from 266 United States Supreme Court cases,⁴⁴ ninety randomly selected state supreme court cases, and a hundred federal appellate cases.⁴⁵ In total, we analyzed 648 court opinions and 882 appellate briefs. Logistic regression⁴⁶ was used to evaluate the impact of readability of the appellant and appellee briefs on the odds of reversal in the state supreme courts and the federal courts of appeals, and then again to evaluate the impact of readability of the petitioner and the respondent briefs at the Supreme Court.⁴⁷ When considering cases from the state

43. See *supra* n. 28 and accompanying text.

44. The 266 Supreme Court cases included the opinions written by the Court and the petitioner and respondent briefs written to the United States Supreme Court from the time Justice Alito first participated in an opinion of the Court on February 21, 2006, through the opinion of the Court issued on June 28, 2007. The database includes only cases in which the Court issued an opinion.

45. We used the same 400 randomly selected cases that we used for our previous intensifier study. To obtain the sample of 200 federal cases, we randomly chose cases from 2001–2003 and randomly selected cases so that the appellate courts hearing the most cases (i.e., the Ninth Circuit and Fifth Circuits) had a proportionally larger representation in the sample. The 200 state cases were randomly chosen in a similar manner, with larger states generally having more cases in the sample. We selected only civil cases that had a clearly discernable outcome, usually either “reversed” or “affirmed,” and the selected cases had at least one brief from each party, usually the principal and the response brief. Occasionally, a reply brief was used if a principal brief was unavailable.

46. Like standard regression analysis, logistic regression is used to model the relationship between a response variable and one or more explanatory variables: It helps a statistician understand the relationship—if any—between the variables considered. As in standard regression, the explanatory variables in logistic regression can be a mix of quantitative and categorical variables. The unique facet of logistic regression is that the response variable is categorical (in the context of this article, this means that a particular case falls clearly into the “reversed” group or the “affirmed” group) instead of quantitative (as would be the case in a study involving variables like “exam score” or “annual salary”). We used logistic regression in our analysis to model the odds of reversal using the explanatory variables (state supreme court or federal court of appeals, for example) that we measured.

47. In evaluating the impact of readability, we initially chose to use the Flesch Reading Ease scale because, given Microsoft Word's practice of rounding the readability measures

supreme courts or federal courts of appeals, we considered the following covariates' impact on the probability of reversal: jurisdiction (federal or state), standard of review (de novo, abuse of discretion, clear error, or other), judicial dissent status (present or absent), and the readability of the court's written opinion. Interactions between these factors were also considered in the statistical model. In the Supreme Court cases, we considered covariates for constitutional status (constitutional issue or not), criminal status (criminal or civil case), judicial dissent status (present or absent), and the readability of the court's written opinion.

Backward elimination⁴⁸ was used to eliminate non-significant factors and interactions from each logistic regression model. In each of the two models (state supreme court or federal court of appeals on the one hand, and United States Supreme Court on the other), the readability of briefs submitted to the court being considered never appeared significantly associated with the outcome of the case. In the model using the state or federal court of appeals data, after backward elimination of non-significant terms, only the jurisdiction variable was a significant predictor of reversal, and this occurred only because state cases have much higher reversal rates than those decided by the federal courts of appeals. In the model for the Supreme Court data, no variables remained in the model after backward elimination, indicating that none of the variables studied was significantly associated with the outcomes of the cases analyzed.

B. The Results

The analysis indicates that the Flesch Reading Ease scores at the state and federal levels are not significantly related to outcome at the five-percent significance level. The same was

to the nearest tenth, Flesch Reading Ease is a more precise measure. All analyses were later run using the Flesch-Kincaid Readability scale with no difference in conclusions.

48. Backward elimination is a form of stepwise regression in which the variables to be included in the model are chosen via an automated process. All of the explanatory variables and interactions are included the first time the model is used, but if any of those variables is shown to be non-significant (which means that it is shown to have a p-value of greater than 0.05) when the model is run, the least significant variable is dropped and the model is refitted. This process is repeated until all non-significant variables are eliminated from the model.

true for the Flesch-Kincaid Grade Level. For the analyses of the Supreme Court data, again there was no statistical evidence that readability or grade level had any impact on outcome.⁴⁹

Although there was no significant relationship between readability and outcome, there were some interesting relationships between the readability of briefs and opinions in the state supreme courts, the federal courts of appeals, and the Supreme Court. For these relationships, we consider statistical significance to be achieved if the p-value for the statistical test is less than 0.00167, which is found by dividing 0.05 by the total number of mean comparisons. (In the context of our research, the calculation is $0.05 \div 30$.)⁵⁰

First, there is no significant difference between the readability of opinions from state supreme courts and federal courts of appeals. They are apparently quite similar in culture with respect to readability of briefs and opinions. Second, United States Supreme Court readability is significantly different from that in the lower courts within all writer comparisons (except in Supreme Court dissents versus those in the federal courts of appeals, where there are comparatively few dissents for analysis). All other within-writer comparisons are significant—United States Supreme Court petitioners, respondents, majority opinion writers, and dissenting opinion writers are all less readable on average than their counterparts in the state supreme courts and federal courts of appeals. Third, while there is virtually no difference among all writers in the state supreme courts, there is some limited evidence of judges writing in a less readable style than lawyers in the federal courts of appeals. Fourth, at the United States Supreme Court, the Justices' writing is significantly less readable than the lawyers'

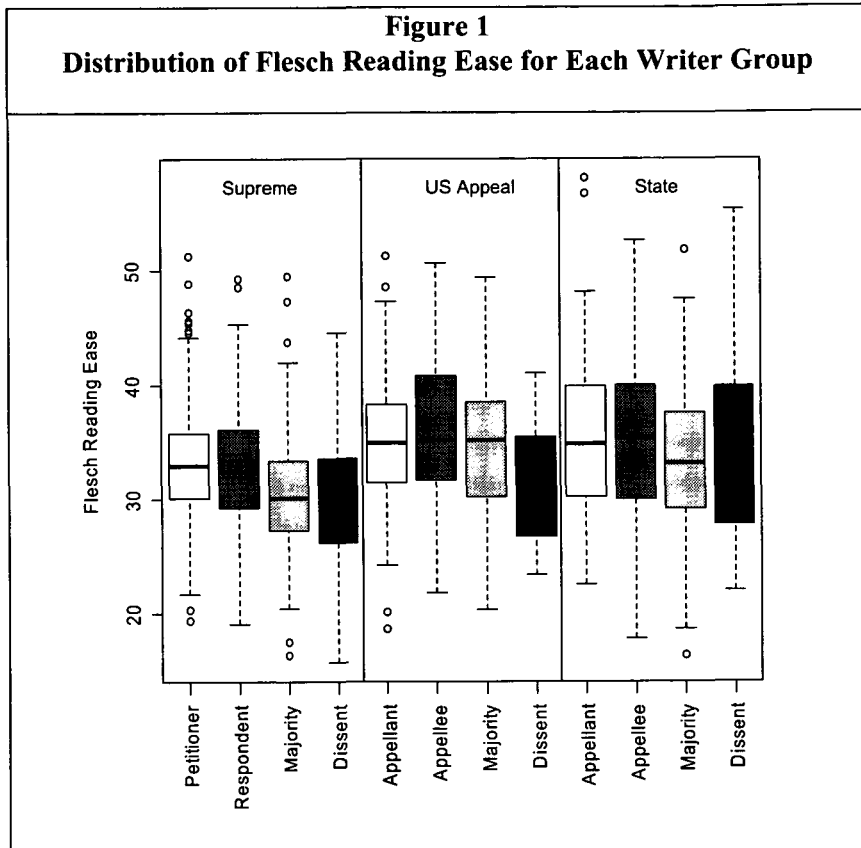
49. The Reading Ease scores and the Grade Level scores differ significantly from those reported in Coleman & Phung, *supra* n. 42. Of course different cases were used in each study, but our scores probably indicate a lower level of readability for Supreme Court briefs and opinions because of a difference in methodology; we included case citations as part of the text, while the Coleman/Phung method deleted citations and inserted the word "scite" in place of each. *Id.* at 81. We included the citations because some legal writing scholars believe that citation within the text renders the text less readable. *See e.g.* Antonin Scalia & Bryan A. Garner, *Making Your Case*, 132–35 (Thomson/West 2008).

50. This approach, known as a Bonferroni adjustment, lessens the possibility of finding spurious significant relationships when testing multiple hypotheses on a single set of data. *See e.g.* Rupert G. Miller, Jr., *Simultaneous Statistical Inference* 6–8 (2d ed., Springer-Verlag 1981).

writing. The differences in the readability and grade-level scores of opinions and briefs in the three sets of courts—the United States Supreme Court, the federal courts of appeals, and the state supreme courts—are shown in Table 1 on this page, and boxplots⁵¹ for readability are given in Figure 1, which appears on the page that follows Table 1. We note in Figure 1 that although the means available to different writers are often significantly different (see the discussion in Section IV), there is still a great deal of variability within each writer group.

Mean Reading-Ease Measures	Supreme	US Appeal	State
Appellee/Petitioner	32.96	34.89	35.14
Appellant/Respondent	32.70	36.27	35.06
Majority (overall)	30.53	34.71	34.04
–unanimous	30.08	35.03	33.50
–split vote	31.16	—**	34.93
Dissent	29.50	—**	36.12
Mean Grade-Level Measures	Supreme	US Appeal	State
Appellee/Petitioner	13.71	13.41	13.48
Appellant/Respondent	13.73	12.93	13.60
Majority (overall)	14.31	13.59	13.68
–unanimous	14.38	13.53	13.89
–split vote	14.21	—**	13.33
Dissent	14.78	—**	13.35
*Higher reading-ease scores and lower grade-level scores mean that the writing should be more readable.			
**The sample size is too small to provide reliable numbers for these categories.			

51. A boxplot illustrates the distribution of an observed variable. The lower and upper ends of the box denote the twenty-fifth and seventy-fifth percentiles of the variable's distribution, with the line in the middle of the box denoting the median (i.e., the middle observation). The whiskers extending from the lower and upper ends of the box denote the observations in the lowest and highest quartiles of the data; the circles appearing beyond the ends of the whiskers denote unusually large or small values (which are typically referred to in statistical analysis as "outliers").



IV. DISCUSSION AND CONCLUSION

The results of our analysis show no correlation between outcome on appeal and readability based on the Microsoft Word version of the Flesch Reading Ease scale and the Flesch-Kincaid Grade Level scale. This result was expected—although it would have been far more interesting to have found a correlation.⁵² Our results also differ from our previous study, in which we found

52. Our results are consistent with those of Coleman and Phung, *supra* n. 42, in this regard.

that the use of intensifiers is significantly related to outcome on appeal.⁵³

The lack of correlation between outcome and readability could be due to several factors. First, there is not much difference in the readability of all the opinions and briefs studied. The lowest readability is found in dissenting opinions of Supreme Court Justices (29.50 Reading Ease and 14.78 Grade Level). The highest readability was found in the briefs of appellants and petitioners in the federal courts of appeals and the Supreme Court (36.27 Reading Ease and 12.93 Grade Level).

Second, it is possible that at the appellate level, where the arguments are likely more developed and studied (by both the lawyers and judges), the merits of the case outweigh the small differences in readability that do exist. This is, of course, a welcome interpretation since our system of justice emphasizes decisions based on the merits,⁵⁴ and our rules of procedure are written and applied so as to reach a decision based on the merits whenever possible.⁵⁵

Third, it could be simply that since the level of readability is well within the educational level of the reading audience, the small differences do not affect the reader's comprehension of the brief or the reader's perceived credibility of the brief writer. Of course, it could also be a combination of the three above factors or some other unidentified factor. It is interesting, but not statistically relevant, to note that dissenting United States Supreme Court Justices tend to write in a less readable manner when writing a dissent. This examination of individual writing styles of the Supreme Court Justices will appear in a forthcoming paper.

Even though no correlation was found between outcome and readability, there are still helpful lessons to be gleaned from our results. First, if a novice lawyer or judge wants to check whether his or her writing is near the readability level of other

53. Long & Christensen, *supra* n. 6, at 182–89.

54. *See e.g. Foman v. Davis*, 371 U.S. 178, 181–82 (1962) (stating that it is “entirely contrary to the spirit of the Federal Rules of Civil Procedure for decisions on the merits to be avoided on the basis of . . . mere technicalities” and that “the purpose of pleading is to facilitate a proper decision on the merits”).

55. *Id.*; *see also e.g. Krupski v. Costa Crociere S.p.A.*, ___ U.S. ___, ___, 130 S. Ct. 2485, 2494 (2010) (noting that the preference of the Federal Rules of Civil Procedure is to resolve disputes on their merits).

lawyers or judges, the novice can simply perform a Microsoft Word readability check and compare the results with our results. (Include your citations as part of the readability check.) State court and federal appellate practitioners, judges, and justices should aim for a Reading Ease score of approximately 35 or a Grade Level score of somewhere around 13. Supreme Court practitioners should try to be a little less readable; aim for a readability score of around 33 or a Grade Level score closer to 14. New Supreme Court Justices (take note, Justices Sotomayor and Kagan) should strive for yet less readability when writing for the majority, around 30 to 31 on the Reading Ease scale and about 14.4 on the Grade Level scale. And, when writing a dissenting opinion, go for the least readable style: 29.5 on the Reading Ease scale and 15 on the Grade Level scale. And, if citations are not included in the calculations (we included them in our analysis),⁵⁶ your writing may score even higher. Finally, in case you are wondering, this article earns a Reading Ease score of 30.4 and a Grade Level score of 14.1, which puts us squarely in the realm of majority-writing Supreme Court Justices.

Of course, if our study approximates reality, you are probably already writing your appellate briefs near the same level, and in all probability, if you attempted to modify your readability, it would make little difference in the outcome of an appeal.⁵⁷ What Macbeth said about life is probably true for the impact of appellate brief readability on the outcome of an appeal: “sound and fury/Signifying nothing.”⁵⁸

Finally, nothing in our study suggests that legal brief writers should not seek to write shorter, rather than longer, sentences and use shorter, rather than longer, words. Readers, including justices and judges, generally prefer concise writing

56. See note 49 for an explanation of why we included citations.

57. This seems especially likely to be true when one considers that Coleman and Phung analyzed twenty-five years of Supreme Court briefs and found no correlation. See Coleman & Phung, *supra* n. 42, at 94–95 (noting no correlation between variables measured—including whether the writer was on the winning or losing side of the case—and readability scores).

58. William Shakespeare, *Macbeth* act 5, sc. 5, <http://shakespeare.mit.edu/macbeth/macbeth.5.5.html> (accessed Aug. 23, 2011; copy of relevant page on file with Journal of Appellate Practice and Process).

and perceive “readable” writers as more credible.⁵⁹ So, although you may gain some points with the judge, more readable writing will probably not win your appeal.



59. See nn. 38–42, *supra*, and accompanying text.