RDO OLR
1966
REVUE DE DROIT D'OTTAWA
OTTAWA LAW REVIEW

Message Undeliverable: An Empirical Assessment of the (Un)Readability of Canadian Court Decisions

Mike Madden
University of Ottawa, Faculty of Law

PUBLIÉ EN LIGNE | PUBLISHED ONLINE	VOLUME	NO	PAGES
November 2026 | Novembre 2026	57	2	377–420

CITATION RECOMMANDÉE · RECOMMENDED CITATION

FRANÇAIS

Mike Madden, « Message Undeliverable: An Empirical Assessment of the (Un)Readability of Canadian Court Decisions » (2026) 57:2 RD Ottawa 377.

ENGLISH

Mike Madden, "Message Undeliverable: An Empirical Assessment of the (Un)Readability of Canadian Court Decisions" (2026) 57:2 Ottawa L Rev 377.

rdo-olr.org · lawrev@uottawa.ca · 57 Louis-Pasteur, Ottawa ON K1N 6N5

ÉTABLIE EN · EST. 1966

UNIVERSITÉ D'OTTAWA
FACULTÉ DE DROIT
SECTION DE COMMON LAW

UNIVERSITY OF OTTAWA
FACULTY OF LAW
COMMON LAW SECTION

ARTICLE

Message Undeliverable: An Empirical Assessment of the (Un)Readability of Canadian Court Decisions

Mike Madden

THIS ARTICLE REPORTS on two original empirical studies of the quantitative readability of Canadian court decisions, across court levels and across legal subject areas. Judges often say that members of the general public are part of the audience for their decisions. With these types of assertations in mind, the two studies within this article measure the extent to which court decisions are likely understandable to the people of Canada. Specifically, the article applies a new law-specific readability formula to 1092 recently published court decisions (containing approximately 6.5 million words of text) to assess how likely Canadian decisions are to be understood by readers of different education levels. The results of these studies demonstrate that many, and sometimes most, members of the Canadian population do not have the ability to easily understand the language that judges use within their decisions. By calling attention to the dissonance that exists between judicial desires to communicate directly with members of the public, and the language within court decisions that is likely not understandable to large segments of the population, this article strives to motivate judges to think differently about how they might write decisions in order to achieve their objectives.

CET ARTICLE TRAITE de deux études empiriques originales portant sur la lisibilité quantitative des décisions judiciaires canadiennes, tous niveaux de tribunaux et domaines juridiques confondus. Les juges affirment souvent que leurs décisions s'adressent au grand public, partie intégrante de leur auditoire. Compte tenu de ce genre d'affirmations, les deux études présentées dans cet article évaluent dans quelle mesure les décisions judiciaires sont susceptibles d'être comprises par la population canadienne. Plus précisément, l'article applique une nouvelle formule de lisibilité spécifique au domaine juridique à 1092 décisions judiciaires récemment publiées (contenant environ 6,5 millions de mots) afin d'évaluer dans quelle mesure des lecteurs de différents niveaux d'éducation peuvent comprendre les décisions canadiennes. Selon les résultats de ces études, bon nombre de membres de la population canadienne, voire la plupart d'entre eux, ont de la difficulté à comprendre le langage employé par les juges dans leurs décisions. En attirant l'attention sur le décalage qui existe entre la volonté des juges de communiquer directement avec le grand public et le langage utilisé dans les décisions judiciaires, qui n'est probablement pas compréhensible pour de larges segments de la population, cet article vise à inciter les juges à repenser la manière dont ils rédigent leurs décisions afin d'atteindre leurs objectifs.

CONTENTS

Message Undeliverable: An Empirical Assessment of the (Un)Readability of Canadian Court Decisions
Mike Madden

I. Introduction **380**

II. Background **381**

III. Readability Scores Across Different Court Levels **385**
  A. Method: Assessing Readability Scores Across Court Levels **386**
    1. Case Selection **387**
    2. Text Extraction and File Preparation **388**
    3. Linguistic Processing of Files / MADRS Calculations **390**
  B. Descriptive Results: Assessing Readability Scores Across Court Levels **391**
    1. Putting MADRS Values into Context **391**
    2. Overall MADRS Values by Court Level **393**
  C. Discussion and Analysis: Assessing Readability Scores Across Court Levels **395**
    1. The Big Picture: How Readable are Court Decisions? **396**
    2. Assessing the Differences Across Court Levels **400**
    3. Conclusion on Overall Readability Across Court Levels **405**

IV. Readability Scores Across Different Legal Subject Areas **406**
  A. Method: Assessing Readability Scores for Family, Criminal, and Tax Law Decisions **407**
  B. Results, Discussion, and Analysis: Readability Scores for Family, Criminal, and Tax Law Decisions **408**

V. The Relevance of Readability in an AI-World **415**

VI. Concluding Thoughts: Canadian Caselaw's Readability Deficit **418**

Message Undeliverable: An Empirical Assessment of the (Un)Readability of Canadian Court Decisions

*Mike Madden**
UNIVERSITY OF OTTAWA, FACULTY OF LAW

> "[*T*]*he Court's decisions speak directly to the Canadian public. The public has a right to understand how their highest court is grappling with the most pressing legal issues of the day*."[1]
>
> **THE HONOURABLE MAHMUD JAMAL, SUPREME COURT OF CANADA**

> "*The public is an important audience for decisions as well. A well-written decision should make sense to anyone, whether or not they have legal training.*"[2]
>
> **THE HONOURABLE RITU KHULLAR, CHIEF JUSTICE OF ALBERTA**

> "[*D*]*ecisions must avoid the unnecessary use of jargon that make judges and lawyers feel smart but alienate the ordinary reader*."[3]
>
> **THE HONOURABLE SALLY GOMERY, ONTARIO SUPERIOR COURT OF JUSTICE**

* Mike Madden, CD, B.A., M.A., LL.B., LL.M., Ph.D. This research was generously funded by a Bombardier Canada Graduate Scholarship—Doctoral, awarded by the Social Sciences and Humanities Research Council. I thank professors Elizabeth Judge, Suzanne Bouclin, and Wolfgang Alschner for their guidance throughout my doctoral studies. Their input significantly improved this article.

1 Office of the Commissioner for Federal Judicial Affairs Canada, *The Honourable Mahmud Jamal's Questionnaire* (Ottawa: Office of the Commissioner for Federal Judicial Affairs Canada, last modified 17 June 2021) at Question 4 online: <fja-cmf.gc.ca/scc-csc/2021/nominee-candidat-eng.html>.

2 Department of Justice Canada, *The Honourable Justice Ritu Khullar's Questionnaire* (Ottawa: Department of Justice Canada, last modified 19 April 2017) at Question 4 online: <canada.ca/en/department-justice/news/2017/03/the_honourable_justiceritukhullarsquestionnaire0.html>. I was saddened to learn that Chief Justice Khullar recently passed away on September 3, 2026. However, I think her words here reflect her belief that the law should be used to make the word a better, fairer place—and I think her words are therefore very worthy of being republished here, as a way of calling continued attention to them.

3 Department of Justice Canada, *The Honourable Sally A Gomery's Questionnaire* (Ottawa: Department of Justice Canada, last modified 3 November 2017) at Question 4

I. INTRODUCTION

It is now common for judges in Canada to express the idea that adjudicative decisions are a "public good" that ordinary people should be able to personally read and understand.[4] But are Canadian court decisions actually written in ways that are understandable to the people of Canada? The answer to this question is empirically provided within this article, and in large measure, the answer is "no."

This article reports on two original studies that use a new, law-specific quantitative readability formula to compute readability scores for 1092 different Canadian judicial decisions that were published in 2022, consisting of almost 6.5 million words of text. The first study measures readability levels by court level, and finds that readability levels decrease as court level increases—which is to say, inferior court decisions are the most readable, followed by superior, and then appeal court decisions, with Supreme Court of Canada (SCC) decisions being the least readable. The second study measures readability levels by subject area, and finds that family law decisions are more readable than criminal law decisions, which are more readable than tax law decisions.

The results of each study are discussed in detail within the body of this article. However, it is immediately worth noting that the average decision for each court level, and each legal subject area, requires a reader to have at least some post-secondary education to easily understand the decision. In the cases of average decisions from appeal courts or the SCC, and the average tax law decision, this education level rises to somewhere between a completed undergraduate degree and a completed law degree. But only 26.7 percent of the Canadian population has a completed undergraduate degree, and slightly more than 18 percent of the population has not completed high school (or equivalent) levels of education.[5] In other words,

---

online: <canada.ca/en/department-justice/news/2017/11/the_honourable_sallyagomerysquestionnaire.html>.

4 Department of Justice Canada, *The Honourable Sharon Matthews' Questionnaire* (Ottawa: Department of Justice Canada, last modified 26 March 2019) at Question 4 online: <canada.ca/en/department-justice/news/2019/03/the-honourable-sharon-matthews-questionnaire.html>.

5 Statistics Canada, *Highest Level of Education by Census Year: Canada, Provinces and Territories, Census Metropolitan Areas and Census Agglomerations*, Table 98-10-0384-01 (Ottawa: Statistics Canada, 9 December 2022), online: <www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=9810038401> [Stats Canada, *Highest Level of Education*]; Statistics Canada, *High School Completion by Census Year: Canada, Provinces and Territories, Census Metropolitan Areas and Census Agglomerations*, Table 98-10-0385-01 (Ottawa: Statistics Canada,

there appears to be a substantial inconsistency between judges' aspirations about communicating decisions to their audiences, and the reality of how difficult it would be for many (or most) members of these audiences to actually understand the judges' decisions. Using objective linguistic data, this article quantifies readability levels for the studied decisions in order to highlight where, and how often, these decisions fall short of being understandable to large segments of Canada's population.

Section II of this article offers some brief background information about the readability formula that is used in the ensuing studies. Section III describes the article's first study of readability levels across different court levels, including the study's methods, and then presents and analyzes the results. Section IV is about the second study dealing with readability levels across three major areas of law (family, criminal, and tax law), and also discusses the study's methods and results. Section V explores whether unreadable decisions are truly a problem today if artificial intelligence tools can be used to help readers understand complex documents. The article concludes with an overarching assessment of the relatively unreadable language of Canadian caselaw. The conclusion and its supporting data show that there is significant room for improvement in how judges write decisions if they hope to make them understandable to the general public in Canada.

II. BACKGROUND

In a previous article, I reported on the creation of Madden's Adjudicative Decision Readability Score (MADRS), a law-specific readability formula.[6] This formula was created from human expert assessments about how well people of different education levels would understand text extracts from actual Canadian court and tribunal decisions. It assigns readability scores to adjudicative decisions based on the linguistic properties of the decisions. This formula outperforms other well-known general-purpose readability formulas in terms of predicting human experts' readability scores for adjudicative decision texts.[7]

---

4 October 2023), online: <www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=9810038501> [Stats Canada, *High School Completion*].

6 Mike Madden, "How Understandable are Adjudicative Decisions? Introducing and Applying Law's Own Readability Formula" (2026) 30 J Leg Writing Inst 183.

7 *Ibid*.

The formula relies on six linguistic variables to produce a global readability score. Four of the variables measure word-related properties in the text, as described below.

1. **Kuperman Age of Acquisition** (**Content Words**) / **AoA_CW.** This variable measures the age at which individuals would have understood a particular *content* word if someone used it in front of them.[8] Content words are words that tend to carry lexical meaning, and include nouns, verbs, adjectives, and some adverbs.[9]
2. **Medical Research Council—Imageability** (**All Words**) / **MRC_Imag_AW.** This variable measures the imageability of words used in a text. Imageability refers to the tendency for a word to evoke both a lexical and a visual representation in the mind of a reader.[10] For instance, the word "beach" (for which most of us have a strong mental image) has the highest imageability score in the relevant database, while the word "plenipotentiary" has the lowest score.[11]
3. **Medical Research Council—Concreteness** (**Function Words**) / **MRC_Concr_FW.** This variable measures the concreteness of function words used in a text. Function words include pronouns, articles, conjunctions, and prepositions that fulfil grammatical functions, but

8 Victor Kuperman, Hans Stadthagen-Gonzalez & Marc Brysbaert, "Age-of-Acquisition Ratings for 30,000 English Words" (2012) 44:4 Behavior Research Methods 978 at 980.

9 Helen Bird, Sue Franklin & David Howard, "'Little Words'—Not Really: Function and Content Words in Normal and Aphasic Speech" (2002) 15:3/4/5 J Neurolinguistics 209 at 210.

10 Maya M Khanna & Michael J Cortese, "How Well Imageability, Concreteness, Perceptual Strength, and Action Strength Predict Recognition Memory, Lexical Decision, and Reading Aloud Performance" (2021) 29:5 Memory 622 at 622. Imageability scores are drawn from the scores that were assigned to 9,240 different words found in Coltheart's Medical Research Council Psycholinguistic database (see Max Coltheart, "The MRC Psycholinguistic Database" (1981) 33:4 QJ Experimental Psychology 497).

11 University of Western Australia, School of Psychology, "MRC Machine Usable Dictionary, Version 2.00" (6 March 1997), online (database): <websites.psychology.uwa.edu.au/school/MRCDatabase/mrc2.html>. Although the machine searchable database is no longer active, my search was performed when it was active and I last visited the database on 12 October 2025. Using a publicly available database containing all the words in the MRC database and all their scores, I downloaded the .CSV file and "sorted" by "imageability" (high to low), and found "beach" has the highest score of 667, and "plenipotentiary" has the lowest score of 129 (see Stephan Akkerman, "MRC-psycholinguistic-database" (last accessed 30 September 2026), online (database): <huggingface.co/datasets/StephanAkkerman/MRC-psycholinguistic-database/blob/main/mrc_psycholinguistic_database.csv> [Akkerman, MRC Psycholinguistic Database]).

that do not tend to carry much lexical meaning.[12] Concreteness means to the extent to which something can be experienced by the senses.[13] For instance, the word "me" (a pronoun for a person whose existence can be perceived through touch, sight, smell, hearing, and taste) has the highest concreteness score in the relevant database, while the function word "as" has the lowest score.[14]

4. **General Inquirer Database—Legal** (**Nouns**) / **GI_Leg_Nouns.** This variable measures the proportion of words in a text that are found on a defined list of 192 words associated with legal, judicial, or police matters.[15] Examples of nouns contained within the General Inquirer's legal word list (*GI_Leg_Nouns*) include the following: advocate, auditor, deposition, fugitive, junta, and legislator.

Another of the MADRS variables measures how commonly researchers can find three-word phrases within a text in a database of frequently used phrases.

5. **COCA—Spoken** (**Proportion of Trigrams in the top 30K**) / **COCA_Spoken_TriTop30K.** This variable measures how often trigrams (three-word phrases) within a text can be found within a list of the 30,000 most-frequently used trigrams contained inside of the Corpus of Contemporary American English—Spoken (*COCA—Spoken*) sub-corpus. Individuals process frequently used multi-word phrases more easily than uncommon multi-word phrases.[16]

---

12 Sidney J Segalowitz & Korri C Lane, "Lexical Access of Function Versus Content Words" (2000) 75:3 Brain & Language 376 at 376.

13 Khanna & Cortese, *supra* note 10 at 623–24. Concreteness scores are drawn from the scores that were assigned to 8,228 different words found in Coltheart's Medical Research Council Psycholinguistic database (see Coltheart, *supra* note 10).

14 Akkerman, MRC Psycholinguistic Database, *supra* note 11. Although the file does not specifically label words as function words, I can confirm from my review of the sorted CSV file that "me" has the highest concreteness score (511), and "as" has the lowest (158).

15 Harvard University General Inquirer, "Descriptions of Inquirer Categories and Use of Inquirer Dictionaries" (12 September 2002), online: <inquirer.sites.fas.harvard.edu/homecat.htm>. The "legal" word list originates from the Harvard IV-4 dictionary lists used by the General Inquirer, a 1966 database consisting of 11,000 words assigned to 119 different word lists (see Scott A Crossley, Kristopher Kyle & Danielle S McNamara, "Sentiment Analysis and Social Cognition Engine (SEANCE): An Automatic Tool for Sentiment, Social Cognition, and Social Order Analysis" (2017) 49:3 Behavior Research Methods 803 at 807).

16 See generally Inbal Arnon & Neal Snider, "More Than Words: Frequency Effects for Multi-Word Phrases" (2010) 62:1 J Memory & Language 67.

The final MADRS variable measures an element of the grammatical complexity of phrases within a text.

6. **Dependents / nominal (no pronouns, standard deviation) / Nominal_Deps_NN_StDev.** The variable measures the number of dependent clauses associated with each noun-phrase in a text, and computes the standard deviation for this number.[17] A higher score for the variable *Nominal_Deps_NN_StDev* would likely be reflective of a text that has relatively more sentences with multiple dependents per nominal. In other words, a higher score would generally mean that a text uses longer sentences with deeper syntactical structures.

The six variables that contribute to the MADRS formula are weighted and combined with a constant to produce a readability score, according to the following statistically derived equation:

$$\begin{aligned}\textbf{MADRS} = {} & 10.036 \\ & + (0.471 \times \textbf{AoA_CW}) \\ & - (0.023 \times \textbf{MRC_Imag_AW}) \\ & + (4.513 \times \textbf{GI_Leg_Nouns}) \\ & - (0.650 \times \textbf{COCA_Spoken_TriTop30K}) \\ & + (1.802 \times \textbf{Nominal_Deps_NN_StDev}) \\ & - (0.018 \times \textbf{MRC_Concr_FW})\end{aligned}$$

MADRS values that this equation produces fall somewhere on a scale from one to seven, based on the scoring criteria contained below in Table 1.

TABLE 1—MADRS SCORING SCALE

This text could be easily understood by an adult who is fluent in English, who has some motivation to read the text, and:	
who has both a law degree and familiarity with the legal subject of the text.	7
who has a law degree.	6
who has an undergraduate degree, or higher level of non-legal education.	5
who has some post-secondary education.	4
who has completed secondary school.	3
who has some secondary school education.	2
who has no secondary school education, or who is less than fluent in English.	1

17 Seymour Chatman, "Pre-Adjectivals in the English Nominal Phrase" (1960) 35:2 American Speech 83 at 83.

The study that initially reports on the readability formula provides much more information about the MADRS formula.[18]

III. READABILITY SCORES ACROSS DIFFERENT COURT LEVELS

The above background information about the MADRS formula will ideally give readers some understanding of, and confidence in, the formula for the purposes of the two studies that follow within this article. The first study, discussed in this section, compares readability levels for decisions from Canadian courts at different levels.

There are compelling reasons why it might be desirable for court decisions to be more readable when they come from higher levels of courts. Decisions of the SCC, for instance, deal with cases that emerge from all parts of the country (that is, they have wide geographical scope), and they are binding on all other Canadian courts as a matter of *stare decisis* (that is, they have the broadest and strongest precedential force of any Canadian court decisions).[19] Furthermore, because the court generally grants leave to appeal only in cases that raise questions of public importance, Canada's apex court tends to deal with cases that should be of greater interest to the Canadian population than cases from other courts.[20] In this sense, the world of potential audience members for SCC decisions is likely to be larger than for a lower court's decisions, so the court may need to express itself in a more readable way if it hopes to reach many of these audience members.

There are also compelling reasons why courts at lower levels may want to produce more readable decisions. For instance, self-represented litigants (SRLs) appear far more often at trial courts than at appeal courts and the SCC.[21] Because of this presence of SRLs, trial court judges may have a stronger obligation to communicate their decisions readably so that the parties can understand the decisions.

In reality, however, we have no current (or even historical) data that compares decision readability levels for trial, appeal, and apex courts in Canada. This information would be helpful to us in making a general

18 Madden, *supra* note 6.

19 *Supreme Court Act*, RSC 1985, c S-26, s 35 [*Supreme Court Act*]; *R v Sullivan*, 2022 SCC 19 at para 59.

20 *Supreme Court Act*, *supra* note 19, s 40(1).

21 Julie Macfarlane & Charlotte Sullivan, *Tracking the Trends of the Self-Represented Litigant Phenomenon: Data from the National Self-Represented Litigants Project, 2019-2021* (Windsor: NSRLP, 2021) at 12.

assessment of readability levels for Canadian courts, so the analysis that follows attempts to provide and explain new and relevant data on this topic.

A. Method: Assessing Readability Scores Across Court Levels

To assess readability levels of decisions from different levels of courts, I first compiled separate corpora[22] of decisions for each court level. That is, I generated a database of text files that comprised either the entire population of decisions being studied (in the case of the SCC), or an appropriate number of randomly identified decisions from the relevant court level (for appeal courts, superior courts, and inferior courts), so that I could measure the readability levels of decisions from each level of court.

Since there was a low number of SCC decisions in 2022, I used the full population of cases within my corpus, subject to the exclusion criteria described below. For all other levels of court, it was not necessary for me to study the whole population of decisions, since the population was large enough that I could reliably study a random sample of the population (of a particular size), with the statistical expectation that this sample would be representative of the larger population. This technique of sampling from the population significantly reduced the time and processing effort required to collect the necessary data.[23] And, for most purposes, it is not necessary to have data about each individual element of a population in order to make use of the data because, if one uses a random sampling method, it is possible to estimate statistically the relevant characteristics of the population as a whole.[24]

In determining the appropriate sample size for each level of court other than the SCC, I used a 95 percent confidence interval, and a seven percent margin of error. Using these two parameters, and the number of total cases within the population of 2022 decisions from each level of court, I then calculated the required sample size using a free online statistical calculator.[25] My random sample sizes were N=183 (appeal court), N=193 (superior court), and N=174 (inferior court). Based on these parameters and sample

22 The words "corpus" and "corpora" are terms that linguists use to refer, respectively, to a database or several databases of language that form the basis of a study.

23 Patrick Dattalo, *Determining Sample Size: Balancing Power, Precision, and Practicality* (New York: Oxford University Press, 2008) at 7.

24 Statistics Canada, *Statistics: Power from Data!*, Catalogue No 12-004-X (Ottawa: Statistics Canada, 2 September 2021) s 3.2.1.

25 The total populations consisted of 2709 appeal court decisions, 9665 superior court decisions, and 1490 inferior court decisions, from which my smaller random samples were

sizes, it is statistically expected that, 19 times out of 20, the true average readability score for the entire population of cases from which my random samples are drawn will fall within +/- seven percent of the sample's average readability score.

1. *Case Selection*

For appeal, superior, and inferior court searches within CanLII (canlii.org), I began with a full text search of the database for documents containing the word "a"—a word that I expect would appear in all cases. I then filtered the search period to only results from 2022, and to only "cases" (excluding secondary sources and legislation). For appeal courts, I used CanLII's filter "all appeal courts" to narrow the search. For both inferior and superior courts, I used the manual filter "select specific courts" to include the relevant courts.

Decisions from Quebec courts were excluded. These decisions were overwhelmingly written in French, and the MADRS readability formula is based on the English language. The variables that contribute to the MADRS formula all rely on measurements about uniquely English words, phrases, or grammatical structures that would not be found in French decisions. It is likely for these kinds of reasons that scholars who have explored cross-language readability formulas conclude that such formulas are unlikely to exist.[26]

For a similar reason, I made best efforts to exclude any cases where it was evident that the decision had originally been written in French before being translated to English. Exclusions on this basis often arose for Federal Court decisions (where the text "[ENGLISH TRANSLATION]" would appear at the top of certain decisions that had originally been written in French). Because translated decisions might not express the authoring judge's intentions as accurately as decisions that were written first in English, I excluded these decisions from my corpora where they were identified.

Additionally, any opinion with less than 400 words was excluded. The MADRS formula was developed using texts of approximately 600 words, so opinions with less than 400 words would not necessarily provide enough linguistic information for the MADRS formula to accurately score the

---

respectively drawn. Sarah Fisher, "Sample Size Calculator" (21 March 2023), online (blog): <qualtrics.com/blog/calculating-sample-size/>.

26 Jesse Saba Kirchner, Justin Nuger & Yi Zhang, "An Extensible Crosslinguistic Readability Framework" (Paper delivered at the Proceedings of the 2nd Workshop on Building and Using Comparable Corpora, Singapore, 6 August 2009) at 11 [unpublished].

text.[27] However, as long as a file contains enough linguistic information, the MADRS formula is length-independent: all six variables that contribute to the formula are ratio or proportion variables that measure the *density* of language features within a text, rather than a raw score of these features (that would inevitably skew higher for longer texts).[28]

Each text file contains a single judicial opinion, not necessarily the entire decision. Where a decision other than a SCC decision contained more than one opinion, only the majority or controlling opinion was included. All SCC opinions meeting the criteria were included, since even dissenting and concurring opinions from that court tend to be influential, or to capture more widespread interest.[29]

Cases from the full population of a relevant court's decisions were randomly identified for inclusion using an online randomization tool.[30] In any situation where a randomly selected case did not meet the criteria for inclusion, then the next-in-sequence case that was not already marked for inclusion was selected instead. Case selection details are publicly available for each of the appeal court, superior court, and inferior court corpora.[31]

2. *Text Extraction and File Preparation*

Once a case was identified for inclusion within the corpus for the relevant court level, I navigated to the web-based publication of the decision and

27 Madden, *supra* note 6.

28 Although one might intuitively expect that longer texts would be harder to read than shorter texts, research does not support this intuition. For instance, one study found that there were no statistically significant differences in reading comprehension levels for students who read longer text passages when compared to students who read condensed passages that were 1/3 shorter than the original passages (see Saeed Mehrpour & Abdolmehdi Riazi, "The Impact of Text Length on EFL Students' Reading Comprehension" (2004) 3:6 Asian EFL J 1). Similarly, another paper found that there was no statistically significant difference in reading comprehension scores between subjects who read 600-word passages and those who read 2000-word passages (see Kristen L Beach, *The Effect of Media, Text Length, and Reading Rates on College Student Reading Comprehension Levels* (DBA Dissertation, University of Phoenix, 2008) at 67 [unpublished]).

29 See generally The Honourable Claire L'Heureux-Dubé, "The Dissenting Opinion: Voice of the Future?" (2000) 38:3 Osgoode Hall LJ 495. See also Christina M Frohock, "Schrödinger's Dissent: The Hybrid Authority of a Dissenting Opinion" (2024) 107:4 Marq L Rev 963.

30 Research Randomizer, "Research Randomizer" (last visited 26 September 2026) online: <randomizer.org>.

31 Mike Madden, "MADRS_MaddenDissertation_Chapter4_AppealCourtSampled+Population" (2025), online (dataset): <doi.org/10.7910/DVN/HPFJCP>; Mike Madden, "MADRS_MaddenDissertation_Chapter4_SupCourtSampled+Population" (2025), online (dataset): <doi.org/10.7910/DVN/QAHP2R>; Mike Madden, "MADRS_MaddenDissertation_Chapter4_InfCourtSampled+Population" (2025), online (dataset): <doi.org/10.7910/DVN/F41BKZ>.

manually copied from the first word to the last word of the body of the decision (or opinion, as appropriate).[32] As this method choice suggests, all header and preliminary information that preceded the body of a decision (such as headnotes, counsel names, hearing location and date, docket number, *etc*.) and all end-matter that followed the last word of the body of a decision (such as endnotes, counsel names, court orders that did not form part of the body of the decision, *etc*.) were omitted from my selections of text. The selected text was then pasted into a new plain text file and saved within a file directory specific for that court level.

This manual approach was time-consuming, but reliable in terms of consistently selecting and extracting only the body of decision for study. My preliminary attempts to draw cases from existing bulk caselaw datasets was unsuccessful, because the full texts in these datasets included all front- and end-matter associated with a decision.[33] Special characters were also not consistently represented within these datasets in the same ways as they appeared within the original decisions. In light of these initial challenges, I decided against using automated approaches for building text corpora. My confidence that such approaches could be used without introducing uncertainty and error into linguistic readability calculations was not high enough to justify the savings in time that the approaches would have offered me.

Although I removed all endnotes from the decisions, I did not remove any inline or in-text citations from the files—even though these text features often contain the same kinds of information as endnotes. Unlike endnote text (which a reader has the choice to read by looking at the end of a document), an inline citation appears within the body of a decision. A reader therefore has no choice but to encounter inline citations when reading and trying to understand a decision. Additionally, inline citations were included within the original decision text extracts that were used to

32 I assessed that this method was consistent with CanLII's Terms of Use and with the organization's prohibition on bulk and systematic downloads (see CanLII, "Terms of Use" (last modified 3 June 2026) s 5(1)(c), online: <canlii.org/info/terms.html>). My downloading of decision texts was very limited in the context of the size of the CanLII database, and it was targeted for research purposes. Additionally, I had discussed this project during its conceptual stages in 2022 with the then-CEO of CanLII, who had no objections to my method.

33 Refugee Law Lab, "Bulk Legal Datasets" (last visited 26 September 2026), online <refugeelab.ca/bulk-data/>; Sean Rehaag & Simon Wallace, "A2AJ Canadian Legal Data" (last modified August 2025), online (dataset): <a2aj.ca/canadian-legal-data/>.

create the MADRS formula, so their impacts on readability are theoretically accounted for within the formula.[34]

3. *Linguistic Processing of Files / MADRS Calculations*

I processed all text files containing court decisions/opinions through three different linguistic software applications to measure the relevant variables within the files that contribute to a text's MADRS value. The *Tool for the Automatic Analysis of Lexical Sophistication* (TAALES) program computes scores for the following MADRS variables: *MRC_Imag_AW*; *MRC_Conc_FW*; *AoA_CW*; and *COCA_Spoken_TriTop30K*.[35] The *Sentiment Analysis and Cognition Engine* (SEANCE) program computes scores for the MADRS variable *GI_Leg_Nouns*.[36] And the *Tool for the Automatic Analysis of Syntactic Sophistication and Complexity* (TAASSC) program computes scores for the MADRS variable *Nominal_Deps_NN_StDev*.[37] Workbooks containing the output from calculations performed by each linguistic software tool for the relevant corpora are publicly available.[38] These measurements were then used to compute MADRS results for each file.

---

34 Mike Madden, "MADRS_MaddenDissertation_Chapter4_SCCLinguisticResults" (2025), online (dataset): <doi.org/10.7910/DVN/XVUB91>; Mike Madden, "MADRS_MaddenDissertation_Chapter4_AppealCourtLinguisticResults" (2025), online (dataset): <doi.org/10.7910/DVN/7C2FTG; Mike Madden, "MADRS_MaddenDissertation_Chapter4_SupCourtLinguisticResults" (2025), online (dataset): <doi.org/10.7910/DVN/BQE66O>; Mike Madden, "MADRS_MaddenDissertation_Chapter4_InfCourtLinguisticResults" (2025), online (dataset): <doi.org/10.7910/DVN/WLR4GR>.

35 Kristopher Kyle, Scott Crosley & Cynthia Berger, "The Tool for the Automatic Analysis of Lexical Sophistication (TAALES): Version 2.0" (2018) 50:3 Behavior Research Methods 1030.

36 Crossley, Kyle & McNamara, *supra* note 15.

37 Kristopher Kyle, *Measuring Syntactic Development in L2 Writing: Fine Grained Indices of Syntactic Complexity and Usage-Based Indices of Syntactic Sophistication* (PhD Dissertation, Georgia State University, 2016) [unpublished].

38 Mike Madden, "MADRS_MaddenDissertation_Chapter4_SCCLinguisticResults" (2025), online (dataset): <doi.org/10.7910/DVN/XVUB91>; Mike Madden, "MADRS_MaddenDissertation_Chapter4_AppealCourtLinguisticResults" (2025), online (dataset): <doi.org/10.7910/DVN/7C2FTG>; Mike Madden, "MADRS_MaddenDissertation_Chapter4_SupCourtLinguisticResults" (2025), online (dataset): <doi.org/10.7910/DVN/BQE66O>; Mike Madden, "MADRS_MaddenDissertation_Chapter4_InfCourtLinguisticResults" (2025), online (dataset): <doi.org/10.7910/DVN/WLR4GR>.

B. Descriptive Results: Assessing Readability Scores Across Court Levels

For the purposes of assessing how readability levels might differ between courts of different levels in Canada, I ultimately calculated readability scores from four different corpora—one corpus each for decisions from the SCC, appeal courts, superior courts, and inferior courts. Relevant characteristics of each corpus are shown, below, in Table 2.

TABLE 2—PROPERTIES OF EACH COURT-LEVEL CORPUS

Court Level	Number of Decision/ Opinion Files	Number of Words of Text (nearest 1000)	Average Decision Length (# of Words)
Apex	75	633,000	8440
Appeal[39]	183	889,000	4857
Superior[40]	193	1,125,000	5829
Inferior[41]	174	1,050,000	6034
Total	625	3,697,000	5915

Spreadsheets of the MADRS results and component variable scores for all the SCC, appeal, superior, and inferior court opinions/decisions within these corpora are publicly available.[42]

1. *Putting MADRS Values into Context*

Before presenting the readability results for these different levels of courts, it might be helpful to put MADRS scores into a more familiar context. I therefore calculated MADRS values for two other judicial texts that might be known to most common law lawyers.

39 This corpus is available here: Mike Madden, "MADRS_MaddenDissertation_Chapter4_CorpusAppealOpinions" (2024), online (dataset): <doi.org/10.7910/DVN/WC630E>.

40 This corpus is available here: Mike Madden, "MADRS_MaddenDissertation_Chapter4_CorpusSuperiorOpinions" (2024), online (dataset): <doi.org/10.7910/DVN/KELQZC>.

41 This corpus is available here: Mike Madden, "MADRS_MaddenDissertation_Chapter4_CorpusInferiorOpinions" (2024), online (dataset): <doi.org/10.7910/DVN/XHLKQ2>.

42 Mike Madden, "MADRS_MaddenDissertation_Chapter4_SCCMADRSResults" (2024), online (dataset): <doi.org/10.7910/DVN/TI6QT9>; Mike Madden, "MADRS_MaddenDissertation_Chapter4_AppealMADRSResults" (2024), online (dataset): <doi.org/10.7910/DVN/LE8YAJ>; Mike Madden, "MADRS_MaddenDissertation_Chapter4_SuperiorMADRSResults" (2024), online (dataset): <doi.org/10.7910/DVN/1VWWI6>; Mike Madden, "MADRS_MaddenDissertation_Chapter4_InferiorMADRSResults" (2024), online (dataset): <doi.org/10.7910/DVN/3IC6ZW>.

One of these texts is drawn from a famous English case,[43] wherein Lord Denning opens his opinion by describing how delightful village cricket is in the summertime (before going on to deal with the law of torts that applies to that case). I calculated MADRS values for the opening paragraph of Lord Denning's opinion in this decision (in isolation), and for Lord Denning's entire opinion.[44] The first paragraph scores a 2.7, while the full text of Lord Denning's opinion scores a 3.6. So, while a person with incomplete high school education could easily understand the opening paragraph, this person would need more than a completed secondary school diploma to understand the opinion as a whole. The difference in these results is not surprising: as the opinion moves from general matters of everyday life (at least, for villagers in England in 1977) to more technical legal matters of nuisance, pleadings, easements, prescriptions, and discretionary remedies, one might expect the reading difficulty of the text to increase.

I also examined a Canadian example of a judicial author—Justice David Watt, formerly of the Court of Appeal for Ontario, now retired—who "is seen by some as part of a welcome trend away from impenetrable, legal jargon."[45] I calculated MADRS values for one of Justice Watt's notable decisions in *R v Shafia*, and again produced scores for the opening section of the decision in isolation, and for the decision as a whole.[46] The opening section scored 2.5, while the entire decision scored 5.1. Again, the scores are not surprising when one reads the decision. The opening section is written

43 *Miller v Jackson*, [1977] EWCA Civ 6.

44 *Ibid* ("[i]n summertime village cricket is the delight of everyone. Nearly every village has its own cricket field where the young men play and the old men watch. In the village of Lintz in County Durham they have their own ground, where they have played these last seventy years. They tend it well. The wicket area is well rolled and mown. The outfield is kept short. It has a good club-house for the players and seats for the onlookers" at para 1).

45 Kirk Makin, "The Judge Who Writes Like a Paperback Novelist", *The Globe and Mail* (11 March 2011) A7.

46 2016 ONCA 812. For Justice Watt's opening section, see paras 1–4, reproduced below:

> "Boaters who travel the Rideau Canal system between Colonel By Lake and Lake Ontario pass through a series of locks at Kingston Mills. Lock gates open. Boats enter. Lock gates close. Boats leave.
>
> But not always.
>
> One morning—June 30, 2009—boaters could not enter Kingston Mills Locks. Oil on the water of the upper lock. A closer look revealed its source. Something in the water.
>
> In the water was a Nissan Sentra. Driver's window open. Ignition off, but not locked. Headlights off. Seatbelts unfastened. Front seats reclined. Rear name plate damaged."

in a distinctly different (more conversational, more informal) style than the remainder of the decision. And the vocabulary in the opening section (discussing boats, a car, death, family members) is simple and concrete in comparison with the vocabulary in other parts of the decision (discussing things like the balance of probative value and prejudicial effect, and antemortem statements of the deceased that were not tendered under the state of mind exception to the hearsay rule).

Ideally, my discussion of these examples and their MADRS values will help to illustrate through how the formula applies to the writings of different judges, whose readability scores cover a spectrum of over 2.5 points across the middle of the seven-point MADRS scale.

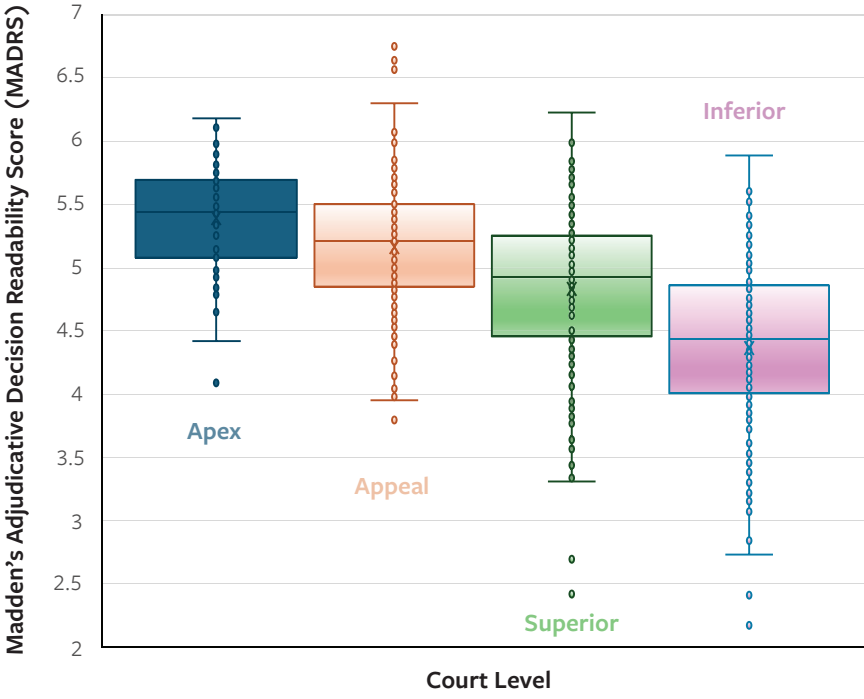
2. *Overall MADRS Values by Court Level*

A workbook containing the full results for decisions in each corpus (separated by tabs), including component scores for each contributing variable, is publicly available.[47] The results are shown below in Chart 1, in the form of boxplots. For each court level, the "box" represents the interquartile range[48] for MADRS values at that court level. The middle horizontal line represents the median score. The "x" represents the average score. The "whiskers" extending above and below each box extend to show the full range of the scores in the jurisdiction, or to 1.5 times the size of the interquartile range—whichever is lesser. Individual data points extending above or below the whiskers represent outlier values. Similar boxplots are shown elsewhere in this article, and should be read in the same manner as the current boxplot.

47 Mike Madden, "MADRS_MaddenDissertation_Chapter4_CourtLevelResults" (2025), online (dataset): <doi.org/10.7910/DVN/CX028J>.

48 "Interquartile range" refers to the range between the 25th and 75th percentiles; it offers a view of the middle 50 percent of the data points, and is less sensitive to outlier points than other dispersion measures (see Michael O Finkelstein & Bruce Levin, *Statistics for Lawyers*, 3rd ed by Stephen E Fienberg (New York: Springer, 2015) at 25).

CHART 1—MADRS VALUES BY COURT LEVEL



As this chart shows, decisions decrease in readability (that is, they have higher MADRS values) for courts that are higher in the judicial hierarchy—with the SCC having the least readable decisions, and inferior courts having the most readable decisions. The average scores for each level of court were the following: Apex=5.40; Appeal=5.16; Superior=4.84; and, Inferior=4.37. The Apex court database of decisions included all opinion types from the court (dissenting, concurring, majority, and unanimous), whereas only majority or unanimous (*i.e.*, controlling) opinions were assessed for all other court levels. This difference in corpus design for the different courts' corpora of decisions could have skewed the above results.

However, the skew appears to have been minimal and does not change the relative readability hierarchy that is shown above. If only the controlling SCC opinions were considered, the results would be almost the same—but with an average of 5.37 for the SCC, instead of the 5.40 that is represented in the above chart.

The most readable decision from any of these courts was *R v SS*,[49] which scored 2.17. This was a criminal case (involving five assault-related charges) that was written by Justice Peter West. The decision spanned 83 paragraphs and consisted of over 12,000 words. Another three of the next four most readable cases were also criminal cases, while one was a family law case. These cases scored 2.41, 2.41, 2.42, and 2.42, and were from the Ontario Court of Justice (again, a decision of Justice Peter West),[50] New Brunswick Court of King's Bench,[51] Provincial Court of British Columbia,[52] and Ontario Court of Justice,[53] respectively.

The least readable decision was *Miller v Canada (National Revenue)*,[54] which scored 6.75. The decision was a tax appeal case, written for a unanimous court. It was 86 paragraphs (11,129 words) long. This decision quotes extensively from a bilingual statute (far more than one would expect from a typical decision), so it is possible that the inclusion of such extensive statutory language, and French language, within the decision, may have contributed to the readability score. The second-least readable decision was a criminal sentencing appeal from the Court of Appeal for Ontario,[55] that was collectively authored by three judges, and consisted of only 755 words. Despite the short length of the decision, it contains a high density of language that is more particular to the legal profession than to everyday English.[56]

C. Discussion and Analysis: Assessing Readability Scores Across Court Levels

The above results offer a useful picture of the general readability levels of Canadian court decisions, broken down in accordance with the levels of courts that issue the decisions. The results also raise both questions and observations that are worth exploring.

---

49 2022 ONCJ 443, West J.

50 *R v Boyle*, 2022 ONCJ 177, West J.

51 *RV v KM*, 2022 NBQB 105.

52 *R v MS*, 2022 BCPC 129.

53 *R v Knight*, 2022 ONCJ 211.

54 2022 FCA 183.

55 *R v Martin*, 2022 ONCA 157.

56 See e.g. *ibid* ("[w]hile the appellant acknowledges that the sentencing judge properly adverted to and described the parity principle, he contends that, on the face of the sentences imposed and the purported gulf between them, there is a clear misapplication of that principle" at para 5).

1. *The Big Picture: How Readable are Court Decisions?*

To begin with, there is a strong chance that most Canadian court decisions could not be easily understood by large segments of the Canadian population. Within the 625 decisions/opinions that were studied here from across all levels of courts, only 11 had MADRS values of less than 3.0—which is to say, only 11 could be easily understood by an adult who has less than a completed secondary school education. Census data indicates that 11.6 percent of individuals in Canada between the ages of 25–64 do not have a "high school diploma or equivalency certificate".[57] For these individuals, the overwhelming majority of court decisions could not be easily understood. And, there is no compelling reason to believe that this 11.6 percent of the Canadian population do not find themselves in family court, or criminal court, or in other situations where they may have a need to understand court decisions.

A perhaps even more concerning reality is that none of the decisions or opinions in the present study scored less than 2.0 on the MADRS scale—which means that none of these decisions are likely to be easily understood by people who have no secondary school education, or who are less than fluent in English. Census data indicates that seven percent of the population does not regularly speak French or English at home.[58] Many of these people may simply choose, for any number of reasons, to speak other languages at home to the exclusion of English and French—despite their fluency in the latter languages. The point, however, is that there may be a non-negligible percentage of Canada's population (potentially amounting to millions of people) whose limited fluency in English, apart from any other contributing factor, prevents them from understanding court decisions easily. These people would likely be unable to understand any of the decisions in the present study.

We do not have data about how many of the above people (with less than a completed high school education, or less than fluent abilities in English) actually want to read court decisions, or have tried and failed in their efforts to understand such decisions. However, the above discussion at least suggests that there may be a communication problem here that requires the attention of our courts. Practically speaking, if someone with less than a completed high school education ever wanted to read and

57 Stats Canada, *High School Completion*, *supra* note 5.

58 Statistics Canada, *While English and French Are Still the Main Languages Spoken in Canada, the Country's Linguistic Diversity Continues to Grow*, in *The Daily*, component of Catalogue No. 11-001-X (Ottawa: Statistics Canada, 17 August 2022) at 25.

understand a Canadian court decision, the odds are almost certain that they would be unable to do so. Where judges purport, or hope, to be writing in ways that make their decisions understandable to ordinary Canadians (as their own words suggest that they do), then the above results raise a red flag: potentially millions of people in Canada do not have the educational or literacy skills that appear to be required to understand approximately 98 percent (614 of 625) of the decisions in this study released by courts of various levels in 2022.[59]

Furthermore, 286 of the 625 studied decisions/opinions scored higher than five on the MADRS scale, indicating that just over 45 percent of all the decisions in this sample could only be easily understood by a person with "an undergraduate degree, or higher level of non-legal education." However, census data indicates that only 32.9 percent of the Canadian population between the ages of 25–64 have a "Bachelor's degree or higher".[60] The disconnect between population capacity to understand a text, and the level at which the text is written, is even more clear here: more than two-thirds of the population would be unable to easily understand almost half of Canada's court decisions.

From a practical perspective, it is impossible to determine whether this disconnect creates a problem that needs to be solved. In order to make that determination, one would probably need more information about whether actual readers (that is, people who are actively trying to understand court decisions) are unable to understand what they are reading. If the evidence (that does not currently exist, and that could not easily be obtained) suggested that this were not the case, then perhaps the disconnect is inconsequential. After all, does it really matter if someone who has no interest in reading a court decision cannot actually understand the content of that decision? Perhaps courts are already tailoring their writing to match the needs of their audiences most of the time, such that the 45 percent of court decisions that score above five on the MADRS scale are probably only of interest to people with a "Bachelor's degree or higher" level of education in any case.

---

59 See e.g. Office of the Commissioner for Federal Judicial Affairs Canada, *The Honourable Sheilah Martin's Questionnaire* (Ottawa: Office of the Commissioner for Federal Judicial Affairs, last modified 21 December 2017) at Question 4 online: <fja.gc.ca/scc-csc/2017-SheilahMartin/nominee-candidat-eng.html>. See also Office of the Commissioner for Federal Judicial Affairs Canada, *The Honourable Nicholas Kasirer's Questionnaire* (Ottawa: Office of the Commissioner for Federal Judicial Affairs, last modified 10 July 2019) at Question 4 online: <fja.gc.ca/scc-csc/2019/nominee-candidat-eng.html>.

60 Stats Canada, *Highest Level of Education*, *supra* note 5.

This hypothesis seems tremendously implausible, since legal issues (and particularly public law legal issues) cut across all elements of the population, regardless of one's education level. People without university degrees still pay taxes, get divorced, commit crimes, invent new products that require patents, get fired unjustly from their jobs, and generally live their lives in wide varieties of ways that might require them to interact with Canada's legal and court systems. These people might need to read court cases—either to understand their legal rights and obligations as a starting point (to assess things like whether they need a lawyer, could proceed to court without a lawyer, or have no legal case worth pursuing), or to understand how their situation has been dealt with by the court after the fact (if they find themselves before a court in defence of their legal rights). As Semple has observed on this point, when

> there are no recent decisions explaining the law, parties will be unable to understand and negotiate for their own rights even if they intend and can afford to. Reported decisions create a "shadow of the law" that lets people understand how the law applies to their own situations, which in turn lets them avoid or settle disputes on fair terms.[61]

But people need to understand the decisions in order for them to benefit from this "shadow" case law.

There is also arguably a causation question here that also needs to be considered. Are there large numbers of ordinary Canadians who are uninterested in reading—or who are currently not trying to read—Canadian court decisions because there is no inherent need for them to ever read these decisions? Or is it possible that many Canadians are currently uninterested in reading such decisions because these people are incapable (or believe that they are incapable) of understanding them? Would more Canadians read court decisions if they were written more readably? The answers to these questions would help to illuminate whether—and to what extent—there may be a real readability problem with Canadian court decisions.

Finally, there is a strong normative argument in favour of making almost all court decisions as widely readable as possible. If it is important for the people of Canada to be capable of seeing and understanding the work that their courts are performing as a way of supervising or ensuring the public accountability of the courts, then surely these rights belong to all citizens,

61 Noel Semple, "Tribunals for Access to Justice in Canada" (2025) 103:1 Can Bar Rev 298 at 323–24 [footnote omitted].

and not only the more educated and more literate citizens.[62] And if parties who are affected by court decisions should be capable of understanding the reasons for disposition of their cases by the courts (out of fairness and respect to the parties), then, again, surely this measure of fairness and respect should extend to all parties, not just the ones who read at higher competency levels. This normative argument for readable court decisions is completely distinct from the empirical argument about how readable court decisions ought to be. The normative argument suggests that people have an equal right to be able to understand court decisions, and this right is undermined when it can only be accessed by individuals along a sliding scale, with higher levels of access afforded to those among us who are more educated, and lower levels of access (or, realistically, no meaningful access) for those among us who have much less formal education.

I accept that all normative arguments must be bounded by practical considerations at a certain point, and that there is likely a limit after which efforts to make decisions more readable to more people becomes unhelpful. For instance, there is likely a point after which further efforts to refine language will result in court decisions that are so linguistically simple that they lose important legal specificity, thereby creating harmful uncertainty in the law. Similarly, there is probably a level of time increment after which any additional time spent by an author in editing and revising a document in order to make it more readable becomes disproportionate to the readability gain that could be achieved by the author: the extra time cost would not be worth the benefit of making the document more understandable to a small number of additional people.

The above MADRS results, however, suggest that most courts and judges have probably not yet reached these points of diminishing or negative returns yet. Inferior court judges have demonstrated that it is possible to write decisions that are, on average, more than a full point more readable on the MADRS scale than the SCC's decisions. There could be a variety of explanations for the differences in scores. But the example set by inferior courts provides persuasive evidence that it is possible to perform the function of judging (as these courts do) while still writing decisions that are relatively more readable.

62 Joseph Goldstein, *The Intelligible Constitution: The Supreme Court's Obligation to Maintain the Constitution as Something We the People Can Understand* (New York: Oxford University Press, 1992) at 19.

2. *Assessing the Differences Across Court Levels*

As I noted above, there are good reasons why we might expect or want decisions to be more readable from courts that are higher on the hierarchy, but also good reasons to expect or want the same thing from courts that are lower on the hierarchy. The results contained in Chart 1, above, show a clear decline in MADRS scores (or increase in readability levels) as one moves down the hierarchy from apex, to appeal, to superior, to inferior courts. It is worth considering why decisions are more readable from lower courts, and also whether this tendency for higher courts to write less readable decisions is desirable.

One explanation for the readability score differences may come from the individualized nature of a trial court judge's work, in contrast with the collegial nature of an appeal or apex court judge's work. For decisions whose results are included in Chart 1 above, all of the inferior court decisions were authored by single judges—as one would expect from this level of court where individual judges preside over trials. And, although at least one Canadian superior court (Ontario's Superior Court of Justice) cross-appoints its judges to a sort of *ad hoc* appellate court (Ontario's Divisional Court, which "hears statutory appeals from decisions of provincial administrative tribunals"[63] that are "usually heard and decided by a panel of three judges but may be heard by a single judge in some circumstances"),[64] none of these superior-appellate Divisional Court decisions were included within the sample represented within Chart 1, above. However, all of the apex court decisions within the sample were written by SCC judges who presided as part of panels, with the smallest panel size in the sample being five judges. It is similarly the general practice of all Canadian courts that I have classified as appeal courts to determine decisions on the merits of an appeal in panels of three judges. In other words, more judges tended to be involved in the decision-making process—although not necessarily in the judgement-writing process—at the appeal court level than at the superior and inferior court levels, and even more still were involved at the apex court level.

The number of judges who were involved in a decision may be a relevant readability factor. In a previous study, I measured readability scores (using a general-purpose readability formula) for judicial opinions from apex court judges across different English-speaking countries (including

63 Ontario Superior Court of Justice, *The Superior Court of Justice: Mapping the Way Forward*, 2010–2012 Court Report (Canada: General Printers, 2012) at 10.

64 *Ibid*.

Canada).[65] The results of this previous study demonstrated that there was a low strength and statistically significant correlation ($r=0.133$, $p<0.001$) between reading complexity scores, and the percentage of judges on a panel who agreed with the opinion.[66] In other words, in cases where more judges signed on to opinions, the opinions tended to be less readable than in cases where fewer judges endorsed the opinions. The same study also found that the average readability score for individually-authored opinions ($N=746$) was substantially better than for collectively-authored opinions ($N=193$).[67] Taken together, these two previous results suggest that the process of authoring a decision in isolation—without any co-authors, and without any possibility of attracting other judges to endorse one's opinion—is associated with better readability scores.[68]

Again, however, the association does not signal causation. We cannot know from this research whether judges write more linguistically complex decisions because they know the decisions will be circulated to their peers on a panel for comment or endorsement (*i.e.*, the primary audience for a panel decision becomes a judge's fellow panel members), whether the decisions only become more complex after other judges have offered their comments and a multi-perspective opinion takes shape, or whether the readability differences between isolated-judge and panel-judge opinions are simply different for coincidental or other as-yet-unexplained reasons.

It may also be the case, however, that lower courts develop an ability to write more readably because the legal subject matter of their decisions is more restricted than higher courts. This hypothesis provides a plausible explanation as to why inferior court decisions are generally more readable than higher court decisions: inferior courts deal mostly with criminal law matters, and in some jurisdictions with family law and/or small claims matters.[69] Perhaps judges of these inferior courts become proficient at

65 Mike Madden, "Judging Readability: A Study of Opinions by Apex-Court Judges from Australia, Canada, South Africa, the United Kingdom, and the United States" (2024) 21 Scribes J Leg Writing 15.

66 *Ibid* at 46.

67 *Ibid* at 39.

68 These results are supported elsewhere in extra-judicial commentary from judges. See e.g. Patricia M Wald, "How I Write" (1993) 4 Scribes J Leg Writing 55 ("persistent colleagues have a wearying habit of pressing for the excision of a particular word, phrase, sentence, or paragraph in the middle of an opinion; and often in our desire to get approval and publish, we will comply. The result is a choppy or erratic transition or no transition at all" at 58).

69 This limited jurisdictional scope of inferior courts' authorities was seen in the very narrow class of subjects that were discussed within the corpus of inferior court decisions that formed part of this study.

explaining how they decide their cases in simple terms because the different legal concepts that they need to explain within their decisions are more finite and narrow in comparison to the range of legal concepts that higher courts encounter. Or, perhaps courts with narrower jurisdiction happen to deal only with cases that fall into legal subject areas that are objectively less complex to explain in English: perhaps criminal and family law, for instance, are inherently easier topics for readers to understand than other legal topics.[70]

The problem, however, is that this hypothesis linking narrower jurisdiction to improved readability scores does not explain the readability level differences between superior and appellate courts. These two levels of courts tend to deal with identically broad ranges of legal subject areas within their decisions, but superior courts have noticeably lower MADRS scores than those of appeal courts. The breadth of legal subject areas that are dealt with at any particular court level therefore offers, at best, an incomplete explanation for readability differences across court levels.

Even without comparing the readability results for apex court decisions to those of other courts, the SCC's results are somewhat disappointing. The audience sizes for SCC decisions should reasonably be larger than the audience sizes for decisions from other levels of courts, since SCC decisions apply across all of Canada, and tend to deal with cases of greater public importance. The MADRS average for SCC opinions (5.37 or 5.4, depending on whether dissenting and concurring opinions are excluded or included, respectively) indicates that something between a completed undergraduate degree and a law degree would be required to understand the opinions. And even the most readable opinion from the Court in 2022 was scored as 4.1 (signaling that at least some post-secondary education is needed to understand the decision). Under these circumstances, it seems clear that the SCC is currently incapable of reaching a large portion of Canada's population through its decisions.

This reality is at odds with the descriptions of the intended audiences for SCC decisions that have been identified by recently appointed judges of that Court.[71] Perhaps the aspiring judges who offered their statements

70 This hypothesis will be examined below when family law, tax law, and criminal law MADRS results are contrasted with one another.

71 See e.g. Office of the Commissioner for Federal Judicial Affairs Canada, *The Honourable Mary Moreau's Questionnaire* (Ottawa: Office of the Commissioner for Federal Judicial Affairs Canada, last modified 26 October 2023), online: <fja-cmf.gc.ca/scc-csc/2023/nominee-candidate-eng.html> ("[t]he Supreme Court's decisions also speak to jurists, law students and the general public, who make up not only today's audience, but also

about the audiences for SCC decisions realized after appointment that the job simply does not allow them to write for all of the audiences who they wanted to reach when they expressed their candidacy for the SCC. Or, notwithstanding their statements within their questionnaires that they complete as part of their candidacy—which they now likely know will be made public, and which therefore must be politically palatable to public audiences—the judges have always understood their primary audience to be members of the profession of law: practicing lawyers and other subordinate judges. Perhaps the nod to members of the public, and the parties, is simply lip service.

But this hypothesis is inconsistent with my own (impressionistic, but extensive) experiences interacting with judges across a wide range of circumstances over the last 15 years. My sense is that judges, almost without exception, are people with integrity, strong work ethics, and concern for those who are both immediately and more remotely affected by their decisions. So when I point out that there seems to be a disconnect between judges' descriptions of their intended audiences, and their abilities to actually be understood by many members of these audiences, I attribute the disconnect to a failed objective rather than a false objective. My hope is that through studies like the present one, judges who realize that they may not be reaching their audiences might act to change their writing in ways that increase the likelihood of reaching these audiences in future decisions.

There is no research showing that desire, alone, is enough to improve the linguistic readability of one's written communications. But for those judges willing to take steps to improve their writing, there is research showing that qualitative writing interventions, in the form of training for decision-makers, and readability-checked templates accompanied by training, do help to improve quality or readability measurements.[72] This

tomorrow's" at Question 4). See also Department of Justice Canada, *The Honourable Michelle O'Bonsawin's Questionnaire* (Ottawa: Department of Justice Canada, last modified 15 April 2021), online: <canada.ca/en/department-justice/news/2021/04/the-honourable-michelle-obonsawins-questionnaire0.html> ("a judge must write in language that is clear, concise and easy to understand...The public must be able to easily understand the reasoning process behind the judge's decision and to conclude that the judicial proceeding was fair" at Question 4).

72 Social Security Tribunal of Canada, "An Evaluation of How Easy it is to Read Decisions of the Social Security Tribunal" (last modified 10 January 2024), online: <sst-tss.gc.ca/en/our-work-our-people/evaluation-easy-it-read-decisions-social-security-tribunal> (The authors found that "[t]he median score of all decisions before the most recent language training was 9 out of 17 points. That improved to 11 after training. Before training only 34 [percent] of decisions scored between 11 and 17 points whereas 51 [percent] scored in this upper

research suggests that it is possible to improve the readability of one's writing if one learns more about what constitutes understandable writing. And it should not be difficult to acquire this knowledge. The National Judicial Institute (NJI) and the Canadian Institute for the Administration of Justice (CIAJ) jointly run multi-day programs twice per year that all new federally appointed judges attend within one year of appointment, and these seminars include sessions on judgement writing and effective communication in the courtroom.[73] Such training should help to ensure that new judges are equipped to write readable decisions. For all other judges, the NJI/CIAJ runs a separate program that offers four days of training focused exclusively on decision-writing.[74] In recognition of the fact that some judges will sit for many years (or even decades) after first being appointed, these education programs might offer more experienced judges an opportunity to refamiliarize themselves with best writing practices.

Even though there are signs, above, that some judges may not be writing highly readable decisions, there are also causes for optimism within the results. Justice Peter West produced two decisions that would likely be easily understandable by adults who have not completed secondary school education. These results from Justice West on decisions that had been randomly included within my sample caused me to wonder whether all of his decisions might be more readable than the decisions of other judges, so I calculated MADRS values for all of his reported decisions from 2022.[75] The average MADRS result was 3.96 (with a range of between 2.17 and 5.35). This average was 0.4 points lower than the average readability score for

---

range after training"). See also Kristie B Hadden et al, "Improving Readability of Informed Consents for Research at an Academic Medical Institution" (2017) 1:6 J Clinical & Translational Science 361 at 363–64. In this study, the authors developed and tested an informed consent template form that was written at the fifth grade level, that future investigators could complete for their own purposes. The authors then compared the readability levels of 82 forms that had been written after the intervention, and found that the mean readability score was two reading grades lower than scores for forms written before the intervention. They also found that the "intervention has resulted in a 658 [percent] increase in the percentage of written informed consents at eighth grade readability or below compared with baseline".

73 National Judicial Institute, "Course Descriptions: January to December 2023" (last visited 26 September 2026) at 6, online (pdf): <nji-inm.ca/index.cfm/publications/public-resources/nji-course-descriptions/>.

74 Canadian Institute for the Administration of Justice, "Mastering the Skill of Judgment Writing (2024 Edition)—Joint program with the NJI" (last visited 26 September 2026), online: <ciaj-icaj.ca/en/upcoming-programs/mastering-judgment-writing-2024>.

75 There were 16 reported decisions by Justice West in the CanLII database from 2022.

inferior court judges, and was 1.44 points lower than the average opinion from SCC judges in 2022.

I have no reason to believe that Justice West is writing readable decisions that are wrong in law, or otherwise lacking in some other dimension of writing or judicial decision-making. For instance, none of his 2022 decisions were the subject of appeals to the Court of Appeal for Ontario that are reported in the CanLII database. Similarly, I have no reason to suspect that Justice West is sacrificing productivity in order to make his decisions more readable: his total of 16 reported decisions in 2022 seems to fall in the middle of the range of between 9–22 reported decisions that I observed from a small sample of other judges from the Ontario Court of Justice that I searched for in CanLII. My point is not to confirm that Justice West is as productive as his peers, but simply to note that there appear to be no reasons to question his productivity, or the quality of his judicial decisions on dimensions outside of readability. Justice West therefore appears to be a leader among his peers in terms of writing readable decisions.

I draw attention to the readability scores of Justice West for two reasons. First, these results and his example establish how it is possible to produce judicial decisions that would likely be understandable to almost the entire population of Canada. The evidence that such readability results can be achieved without apparent compromises in different dimensions of judicial decision-making will ideally motivate other judges to strive to achieve similar results in appropriate cases.

Second, the results offer useful quantitative (non-impressionistic) data about a judge who appears to have unlocked some of the keys to writing readable decisions. Perhaps we should be seeking out perspectives from judges like Justice West about how to write readable decisions if we hope to take the next logical step forward in this line of research by determining not just the "who" of readability, but also the "why" and "how." With data about a judge's MADRS results, the task of identifying people who we might want to interview to learn more about their processes and perspectives on decision-writing becomes a little easier.

3. *Conclusion on Overall Readability Across Court Levels*

The results of this study provide, for the first time, a comprehensive look at the readability levels of decisions from different levels of courts in Canada. As a general matter, all levels of courts are producing decisions that are unlikely to be understandable to millions of people in Canada. This is not optimal. But with this data, it may now be possible to contemplate relevant

interventions for judges who want to change the status quo. Judges of the SCC, who tend to have the least readable decisions in the study, arguably have the most work to do on readability fronts. But these same judges could also show important leadership to other Canadian judges by working to produce more linguistically understandable decisions going forward.

IV. READABILITY SCORES ACROSS DIFFERENT LEGAL SUBJECT AREAS

There are many reasons why it might be useful to understand how readable decisions are within different classes of legal subject areas. We know that in cases involving family law, self-represented litigants are estimated to be parties in up to 80 percent of cases.[76] Under these circumstances, one might expect family law decisions to be generally more readable than decisions in other areas of law where parties are more frequently represented by lawyers. Additionally, judges seem to think that criminal cases are of broad interest to the public, and these judges suggest that they write their criminal law decisions in part so that members of the public can understand the decisions—so one might also expect these decisions to be more readable when compared to other legal subject areas.[77]

In other areas of law, however, the expectation for readable decisions might be reduced. In tax law cases, for instance, the legal subject matter may be more complex so as to defy explanations using simple language within judicial decisions.[78] For years, there have been calls for tax simplification in Canada, although "neither federal nor provincial governments

76 Julie Macfarlane, *The National Self-Represented Litigants Project: Identifying and Meeting the Needs of Self-Represented Litigants* (Windsor: National Self-Represented Litigants Project, 2013) at 15.

77 Department of Justice Canada, *The Honourable Warren B. Milman's Questionnaire* (Ottawa: Department of Justice Canada, last modified 3 November 2017), online: <canada.ca/en/department-justice/news/2017/11/the_honourable_warrenbmilmansquestionnaire.html> ("[t]he public has, in that sense, an interest in the disposition of every case that comes before the courts. This is especially true in criminal cases, where the public is also, in effect, the initiating litigant" at Question 4) [emphasis added]. See also Department of Justice Canada, *The Honourable Cynthia Petersen's Questionnaire* (Ottawa: Department of Justice Canada, last modified 30 May 2022), online: <canada.ca/en/department-justice/news/2022/05/the-honourable-cynthia-petersens-questionnaire.html> ("[i]n matters of criminal law…in many cases the public at large will also have a substantial interest in either the outcome and/or reasoning of the court, and may therefore also be part of the primary audience" at Question 4).

78 Tamer Budak, Simon James & Adrian Sawyer, "The Complexity of Tax Simplification: Experiences From Around the World" in Simon James, Adrian Sawyer & Tamer Budak, eds,

have reacted to the various pleas for simplicity—or complaints about complexity."[79] If the legislation that governs a particular subject is complex and difficult to understand, then perhaps the case law about that subject will be similarly difficult to understand.

To assess how Canadian court decisions might differ in their readability levels across various subject areas, I used the MADRS formula to analyze readability levels for samples of court decisions from three legal subject areas: family law, criminal law, and tax law. Below, I describe my methods for conducting this study, and I discuss the readability results.

A. Method: Assessing Readability Scores for Family, Criminal, and Tax Law Decisions

My research method for this study resembled my method for the previous study in this article. I assembled three separate corpora of judicial decisions/opinions (one each for family law, criminal law, and tax law cases) using a random-sampling approach.[80] I again used a 95 percent confidence interval, and a seven percent margin of error. Using these two parameters, and the number of total cases within the population of 2022 decisions from each subject area, I then calculated the required sample size using a free online statistical calculator.[81] My sample sizes were N=177 (family law); N=184 (criminal law); and N=106 (tax law). Statistically, one can infer that results from these sample sizes would be representative of results (within + /- seven percent) for the full population of cases from which the samples are drawn, 19 times out of 20.

For all three subject areas, I began with a full text search of the CanLII database for documents containing the word "a"—a word that I expect would appear in all cases. I then filtered the search period to only results from 2022, and to only "cases", from "all courts". I manually selected all jurisdictions except Quebec. For the family law and tax law cases, I used CanLII's subject filter, set to "family" and "taxation", respectively. For

The Complexity of Tax Simplification: Experiences From Around the World (Houndmills, Basingstoke, Hampshire: Palgrave Macmillan, 2016) 1 at 3.

79 François Vaillancourt & Richard Bird, "Tax Simplification in Canada: A Journey Not Yet Mapped" in Simon James, Adrian Sawyer & Tamer Budak, eds, *The Complexity of Tax Simplification: Experiences From Around the World* (Houndmills, Basingstoke, Hampshire: Palgrave Macmillan, 2016) 70 at 73.

80 Dattalo, *supra* note 23 at 8.

81 The total populations consisted of 1814 family law, 2788 criminal law, and 227 tax law decisions, from which my smaller random samples were respectively drawn. Fisher, *supra* note 25.

criminal law cases, I used CanLII's subject filter, set to "criminal or statutory infractions". However, since I was primarily interested in true criminal cases (*i.e.*, not regulatory offences), I added the phrase "criminal code" as an additional term within my full text search.

All other exclusion criteria that applied to cases (described above) for my first study were also applied here, for all of the same reasons. Randomization, text extraction, and linguistic score calculations were all achieved in the same manner. Following these steps allowed me to compute MADRS values for decisions/opinions from each of the three legal subject areas. Spreadsheets containing the full population list of cases for each subject area, and the randomly selected cases for each subject area, are publicly available.[82]

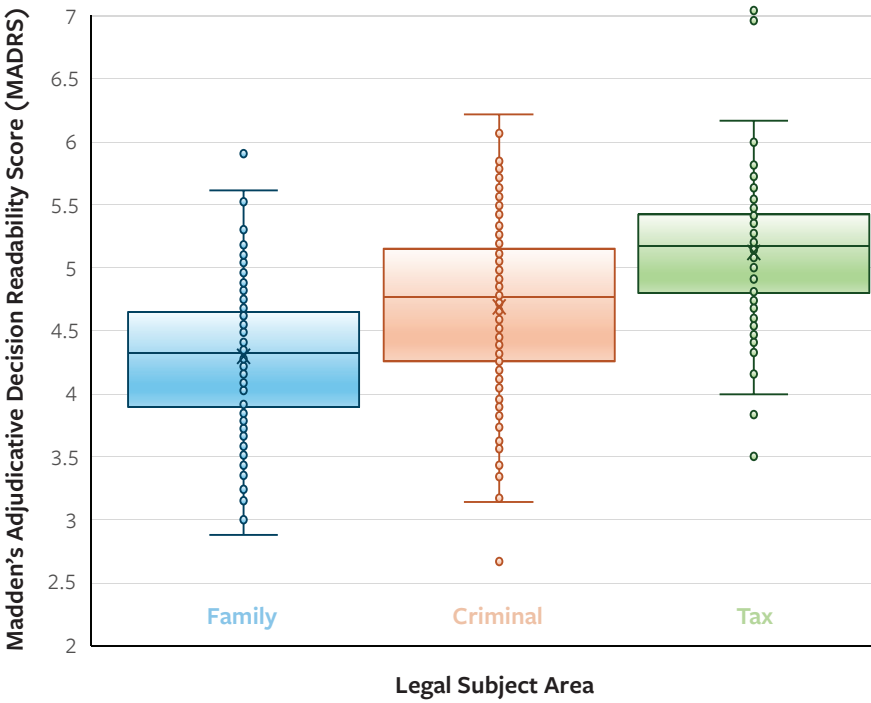
B. Results, Discussion, and Analysis: Readability Scores for Family, Criminal, and Tax Law Decisions

In total, I calculated results for 467 decisions/opinions, containing over 2.8 million words of text. A workbook containing the full results for decisions in each corpus (separated by tabs), including component scores for each contributing variable, is publicly available.[83] A graphical representation of MADRS results by legal subject area is shown below in Chart 2.

82 Mike Madden, "MADRS_MaddenDissertation_Chapter4_CrimLawSampled+Population" (2025), online (dataset): <doi.org/10.7910/DVN/PJPYER>; Mike Madden, "MADRS_MaddenDissertation_Chapter4_FamLawSampled+Population" (2025), online (dataset): <doi.org/10.7910/DVN/MUQPOE>; Mike Madden, "MADRS_MaddenDissertation_Chapter4_TaxLawSampled+Population" (2025), online (dataset): <doi.org/10.7910/DVN/PRKKRM>.

83 Mike Madden, "MADRS_MaddenDissertation_Chapter4_FamilyCrimTaxResults" (2025), online (dataset): <doi.org/10.7910/DVN/N5OIQD>.

CHART 2—MADRS VALUES BY LEGAL SUBJECT AREA



The results on this w show that family law decisions tend to be more readable than criminal law decisions, which in turn are more readable than tax law decisions. The average scores for each subject area were the following: family law=4.30; criminal law=4.69; and, tax law=5.13.

There are two outlier tax law decisions that fall close to the 7.0 level on the MADRS scale.[84] Both of these decisions, however, have values for the variable *Dependents per nominal (no pronouns, standard deviation) / Nominal_Deps_NN_StDev* that are unusually high, and that fall outside of the range for this variable that existed within the files that were used to derive the MADRS formula.[85] The use of regression formulas (like the MADRS formula) on data that falls outside of the scope of the formula's original data is called extrapolation; a regression formula's results may be biased or

84 One outlier case scored a 7.02 (see *Soulliere v Canada*, 2022 FCA 126); the other scored a 6.96 (see *National R&D Inc v Canada*, 2022 FCA 72).

85 These two decisions had values for the variable of 2.91 and 1.96, respectively, whereas the files that were used to derive the MADRS formula had ranges for the variable value of between 0.83 and 1.75.

imprecise when extrapolating.[86] Consequently, MADRS results for these two decisions are not necessarily valid predictions of the scores that expert human readers would assign to the decisions.

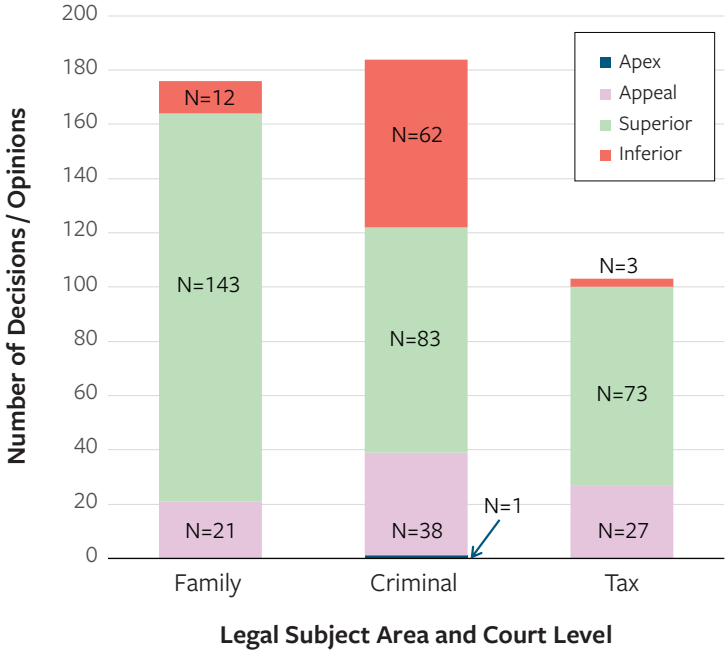
If these two decisions are excluded from the average tax decision readability score, then the average score for decisions within the "tax" legal subject area falls to 5.09—which is still a substantially higher average than the averages for both family and criminal law decisions. Having reviewed the texts of these two outlier decisions, I noticed that they both quoted extensively from legislation that used relatively complex syntactic structures.[87] I attribute the uncharacteristically high scores for the variable *Nominal_Deps_NN_StDev* in these two outlier decisions to the high density of statutory excerpts in each decision.

Although the above chart and MADRS average results show readability levels for family, criminal, and tax law decisions, the cases that contribute to these results were drawn from all levels of court. That is, the process of randomly selecting decisions from the populations of all decisions within each subject area naturally leads to the inclusion of cases from each different level of court that writes decisions on that subject (usually on a proportionate basis). The breakdown of decisions by court level and subject area is shown graphically below in Chart 3.

86 See generally Gerald J Hahn, "The Hazards of Extrapolation in Regression Analysis" (1977) 9:4 J Quality Tech 159.

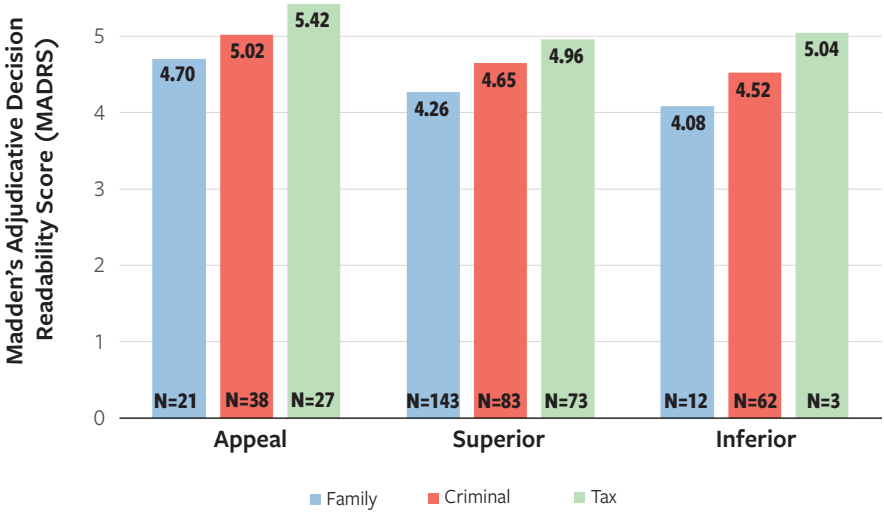
87 Tim R Samples, Katherine Ireland & Caroline Kraczon, "TL;DR: The Law and Linguistics of Social Platform Terms-of-Use" (2024) 39:1 BTLJ 47 ("*syntactic complexity* reflects the range and complexity of language forms in a given text. Compound sentences, embedded structures, and modifying clauses, for instance, make sentences more syntactically complex" at 75 [emphasis in original] [footnote omitted]).

CHART 3—NUMBER OF DECISIONS / OPINIONS BY LEGAL SUBJECT AND COURT LEVEL



Knowing from the previous study that decisions tend to have different readability levels depending on the level of court that produces the decisions, it could be helpful to think about how family, criminal, and tax law decision readability levels vary from one another in general, but also about how they vary from one another at each level of court. In Chart 4, below, the average MADRS value for decisions on each legal subject, at each court level, are shown graphically.

CHART 4—MADRS FAMILY, CRIMINAL, AND TAX AVERAGES BY COURT LEVEL



Two different trends are arguably visible within this chart. First, within each court level, the hierarchy of most readable (lowest score) to least readable (highest score) across subject areas is identical: family law is most readable, then criminal, then tax law. Second, for each individual subject area, the hierarchy of most readable to least readable court level is identical—with one exception. That is, for both family law and criminal law, inferior courts are the most readable, then superior courts, then appeal courts. For tax law, there is a subtle variation of the pattern. Appeal courts are the least readable. But inferior courts occupy the middle ground in tax law cases, where their decisions are slightly less readable than those of superior courts. There was a very small sample size of these inferior court tax law decisions (only three), so it would perhaps be unwise to generalize about the population of inferior court tax law decisions on the basis of this small sample.

The trends that manifest in Chart 4 reinforce my conclusion from the previous study about the relationship between MADRS values and court levels (*i.e.*, that higher MADRS values are associated with higher court levels, and vice-versa), even when controlling for different legal subject areas. And, if one excludes the three inferior court tax law decisions from

consideration, then the trends also support my conclusion that family law decisions are more readable than criminal ones, which are more readable than tax law decisions—even when controlling for the level of court that produces the decisions.

Setting aside how family, criminal, and tax law decisions score from relative perspectives, it is also worth asking whether these classes of decisions are as readable as we might expect or want them to be. Family law decisions are the most readable of any subgroup that has been studied so far. However, 123 of 177 family law decisions scored greater than four on the MADRS scale (indicating that at least some post-secondary education would be needed to understand the decisions), and only one decision scored lower than three on the scale.[88] Since judges seem to be aware of both the high proportion of self-represented litigants who appear in courts on family law matters, and the tendency for self-represented litigants to do their own legal case-law research, the average MADRS value for family law decisions is perhaps unnecessarily high. In other words, while judges seem to be doing a better job of writing at a level that larger groups of people will understand in family law cases, there is still room for improvement.

To be clear, my suggestion here is not that every family law decision needs to be written at an equally easy readability level. A case at a trial court where both parties are highly educated and represented by counsel, where the law is already settled and well-stated within other previous decisions, and where there are no other public interest factors within the case that would attract a wide audience for the court's decision, does not necessarily need to be universally understandable. But it is not always easy to predict in advance when a case will attract only a limited audience, so a judge would be taking a risk of potentially excluding interested people from being able to read a decision if the judge wrongly assumed at the outset that the audience for the decision was narrow.

However, there are likely adults in Canada with only a completed secondary school diploma (or less) who still need to read court decisions about family law (*e.g.*, because they are contemplating divorce, or want greater access to their children after divorce, but they cannot afford to hire a lawyer to represent them). People in this situation would probably need to understand more than just the one out of 177 family law decisions that was actually written at their reading level. Unfortunately, the above results

88 This superior court decision has a MADRS score of 2.88 (see e.g. *Abdulmalik v Cassidy*, 2022 ONSC 4730).

from Chart 2 suggest that there is currently not a critical mass of family law decisions that people on the lower end of the educational or literacy spectrum could look to for guidance, since these decisions are overwhelmingly written at a post-secondary reading level.

The readability results for criminal law cases are also somewhat surprising. Criminal law is another area of law that some judges have specifically identified as being of interest to the broad public, and these judges suggest that they write their criminal law decisions so that the ordinary public can understand the decisions. Furthermore, the offenders in these cases would often be expected to have low levels of education, since "there is a strong negative correlation between educational attainment and various measures of crime",[89] and since a majority of incarcerated offenders in some places have not completed their high school education.[90] Admittedly, most criminally-accused persons in Canada seem to be represented by counsel (who would presumably have a role to play in explaining court decisions to their clients).[91] However, the MADRS average of 4.69 (between some post-secondary and a completed undergraduate degree would be needed to easily understand the decisions) suggests that judges are not necessarily writing their criminal law decisions in ways that could be directly understood by many of the accused persons or offenders who are affected by the decisions.

Perhaps some individuals would be satisfied with having their court decision explained to them by a lawyer—in which case, the readability levels of criminal law decisions would not be a true concern for these people. But the above readability results show that the element of choice has essentially been taken away from these and other individuals with lower literacy levels: regardless of whether they would prefer to read a decision personally, this option will be out of reach for many individuals, since they will be unable to understand the decision if they try to read it without assistance.

The tax law MADRS values are perhaps not surprising, considering how tax law is generally viewed as being complex. And, if one pictures the typical tax law litigant as a wealthy person who is represented by a lawyer, and who has a high level of education—or as a corporation led by sophisticated individuals—then perhaps the above tax law readability scores are

89 L Lochner, "Education and Crime" in Penelope Peterson, Eva Baker & Barry McGaw, eds, *International Encyclopedia of Education*, 3rd ed (Elsevier, 2010) 239 at 241.

90 *Ibid*.

91 Within my corpus of criminal law decisions, there were no cases involving a self-represented litigant.

appropriate. But this is not an accurate picture of every tax law litigant. Consider, for instance, the *Zazula v Canada* (*Attorney General*)[92] decision, which scored 5.27 on the MADRS scale—a score that is higher than the average score for tax law decisions. This decision involved a self-represented litigant who sought judicial review of a Canada Revenue Agency decision to tax the person's accidental over-contributions to his tax-free savings account. In other words, it was a case involving a personal (not corporate) taxation issue, where the applicant was not being assisted by a lawyer. In cases like this, one might expect a court to take greater pains to ensure that its decision can be easily understood by the litigant, and by others who might find themselves in similar situations, instead of writing at a level that requires a person to have somewhere between a completed undergraduate degree and a law degree in order to easily understand the decision.

The above analysis of readability differences across legal subject areas begins to paint a more nuanced picture of the readability landscape for Canadian court decisions. But the analysis reiterates the points that I made earlier when discussing readability scores across court levels: even the most readable subgroup of decisions (family law decisions, in this analysis) is still generally written in ways that make comprehension by large percentages of the population unlikely. Authors of decisions across all these subject areas may wish to consider whether there is more that they can do to improve their decision readability levels.

V. THE RELEVANCE OF READABILITY IN AN AI-WORLD

Before concluding, I must acknowledge that this article was written during a time when generative artificial intelligence (AI) and Natural Language Processing technologies were developing at exceedingly fast rates. Within this context, it is fair to ask whether readability levels of adjudicative decisions matter in a world where machines can summarize these decisions for readers in customizable ways. Are unreadable decisions truly a problem if ordinary members of the public have access to AI tools that can summarize or re-write the decisions for readers in easily understandable language?

This question arguably presumes a more advanced state of publicly and freely available AI technologies than actually exists. In a recent study, Deroy et al tested 15 summarization algorithms against summaries written

92 2022 FC 1156.

by human experts concerning the same 50 cases, to identify how well the summarization tools performed.[93] Among other things, these authors found that the algorithmic summaries tended to suffer from lead-bias (in that they drew disproportionately from the beginning of the summarized text), which meant that they tended to frequently exclude sentences from the "[*r*]*atio of the decision*" part of the summarized text (that typically falls in the middle or end portion of a decision) that human experts included in their summaries.[94] Most AI summaries within this study would therefore be poor substitutes for the original decisions, in terms of explaining the rationales for the decisions—although it would likely be possible for future AI systems to correct for this flaw by drawing on a weighted basis from the relevant sections of a decision in order to summarize the key legal content.

In an even more recent and relevant study, Elliott Ash et al compared participants' objective and subjective comprehension measurements for both AI—and human expert-generated summaries of the Supreme Court of the United States opinions.[95] This study began by noting how most AI language models "do not translate well to the legal domain (as domain shifts are a major obstacle in natural language processing)."[96]

The authors also described at length the general need for "prompt engineering" (that is, the careful and iterative development of prompting questions or instructions that feed into a generative AI system in order to ensure that relevant, responsive, and accurate output is produced by the system) when using generative AI systems, and their own extensive process for creating prompts that would yield the desired summaries of the court opinions.[97] After an initial series of prompts, the summaries generated by the GPT-4 system were substantially less readable than the full court opinions that they summarized.[98] However, after a further prompt that instructed GPT-4 to use simple language within the summaries, the final versions were scored as being "easily [understandable] by 13–15-year-old

93 Aniket Deroy et al, "An Analytical Study of Algorithmic and Expert Summaries of Legal Cases" in Erich Schweighofer, ed, *Legal Knowledge and Information Systems* (Amsterdam: IOS Press BV, 2021) 90 at 91.

94 *Ibid* [emphasis in original].

95 Elliott Ash et al, "Translating Legalese: Enhancing Public Understanding of Court Opinions with Legal Summaries" in *Proceedings of the Third Symposium on Computer Science and Law* (New York: The Association for Computing Machinery, 2024) 136.

96 *Ibid* at 137.

97 *Ibid* at 137–140.

98 *Ibid* at Figure 2.

students."[99] These final summaries were also more readable than ones generated by human experts.[100] This research suggests that AI systems could play a useful role in summarizing adjudicative decisions using language levels that meet the needs of different types of readers.

However, the authors of the study candidly note two main drawbacks that prevent AI summaries from being fully useful to members of the lay public who might need them most. First, "[s]imple prompts such as 'summarize this court opinion' yield unsatisfactory results. Prompt engineering that draws on legal domain knowledge produces summaries that are clear and crisp. Hence, legal experts play a critical role in the development, evaluation, and deployment of AI assistants in legal applications."[101] And, second, "summaries are effective at conveying the main points of a legal argument but can miss some subtle nuances that legal professionals may find important."[102] It is probably unrealistic, then, to expect that every person who wants to understand a court decision will, today, be able to obtain an AI-generated summary that meets that person's needs. These limitations with AI-generated summaries could be significant: if a person cannot figure out how to make a machine compute the necessary answer, or if a machine can only compute an answer that includes some (but not all) of the information that one needs, then the machine is not optimally useful.

For individuals who are interested in Canadian case law summaries and who do not have access to subscription-based databases (like the Westlaw or Lexis databases), the CanLII database now includes AI-generated case summaries for case law from many jurisdictions.[103] However, CanLII also recently announced the launch of a user survey initiative, wherein users provide feedback on the usefulness of these AI-generated summaries.[104] Although there is not yet any published research that assesses the accuracy of CanLII's AI-generated summaries, the existence of this survey seems to suggest that CanLII's summaries are in an early stage of maturation. And, in any event, they are currently not available for many—and in particular,

99 *Ibid* at 141.

100 *Ibid* at Figure 2.

101 *Ibid* at 143.

102 *Ibid.*

103 CanLII, "CanLII and Generative AI" (last modified 29 August 2025), online: <www.canlii.org/info/generative-ai.html> [CanLII, "CanLII and Generative AI"].

104 CanLII, "CanLII Launches Nationwide Survey to Assess AI-Generated Legal Summaries" (1 October 2024), online (blog): <blog.canlii.org/2024/10/01/canlii-launches-nationwide-survey-to-assess-ai-generated-legal-summaries>.

the larger—Canadian jurisdictions like the Federal jurisdiction and Ontario (where only ten percent of decisions have summaries).[105]

Based on the above general limitations, AI summarization and re-writing tools are likely not yet capable of standing in the place of full adjudicative decisions for many of the audiences who might be interested in the cases. Well-crafted summaries certainly have a role to play in many contexts. For instance, law firms who wish to produce summaries for their clients might benefit from these tools, since the firms would likely have enough in-house legal knowledge to engineer useful prompts for the generative AI system, and their clients may be satisfied with high-level details about a case. But the tools are far less suitable today for the parties to a case (who may want to understand the full reasoning behind the decision), and to prospective litigants (who also need to understand nuances in the law in order to understand their own positions).

One could reasonably expect that better AI-generated summaries will soon be available. Improvements in the technology and coverage of these tools is happening quickly. But until they arrive, and until they are priced in such a way as to put them within reach of ordinary members of the public in Canada, the issue of readable judicial decisions remains pressing and relevant.

VI. CONCLUDING THOUGHTS: CANADIAN CASELAW'S READABILITY DEFICIT

Within this article, we have seen the results of two different studies that looked at the readability levels of Canadian court decisions from different perspectives. As a general matter, the results are not encouraging. Across both studies, a total of 1092 decisions or opinions were scored. Not a single decision scored at or below 2.0 on the MADRS scale. And only 12 decisions (approximately 1.1 percent of the studied decisions) scored below 3.0, which is to say, they could be easily understood by a person with less than a completed high school education.

The courts studied in this article may want to write their decisions for the benefit of the Canadian people. And there are many good reasons why these courts would want to reach diverse audiences with their decisions, just as there are many good reasons why the Canadian people might want to read and understand the decisions. But the results discussed in this

105 CanLII, "CanLII and Generative AI", *supra* note 103.

article clearly suggest that an overwhelming proportion of decisions are written in ways that are too difficult for millions of Canadians to understand. There is a disconnect between aspiration and reality that is revealed in this article. The aspiration to write readable decisions is not matched with a practice of writing readable decisions.

I am prepared here to assume that many authors who realize their decisions are likely not reaching their audiences will strive to do something that could change this situation. Since judges' own words show that many of them—across different levels of courts and jurisdictions—see their audiences broadly as including the general public, these same judges probably see the function of communicating in ways that can be understood by the general public as a function that they ought (or would like) to perform. So, it does not strike me as an illogical or unsupportable assumption to conclude that these judges will do something to change the situation if they realize that they are falling short of their own expectations.

Unfortunately, this article doesn't unlock any magic secrets about how to write more readable decisions. The MADRS formula (and the variables that contribute to this formula) suggest that people will easily understand more concrete, imageable, and familiar words, language that avoids technical legal terms, language that aligns with frequently used patterns of speech in English, and short, simple sentence structures. But this is not a revelation. Most judges who have attended decision-writing workshops somewhat recently have probably already received advice to write in ways that should lead to better MADRS results. The challenge for these judges is therefore not so much in knowing, but in doing; it requires additional effort from authors during the decision-writing and editing processes, and perhaps outside help to these authors from linguistic or communications specialists, writing coaches, or other editorial support staff. For judges who have not attended decision-writing training recently, it might be helpful to re-engage with this type of training as a refresher.

If nothing else, however, this article uses empirical data to highlight the current gap that exists between the level at which decisions are generally written, and the (inadequate) level of language and reading competence that many audience members for these decisions have. This article does not fix the problem, but it does take the first step of using quantitative linguistic data to call attention to it in ways that are difficult to ignore. My hope is that judges who really care about reaching their audiences will find motivation within these pages to renew their efforts at communicating readable decisions to the Canadian public.