

2019

## The Impact of Cross-References on the Readability of the U.S. Internal Revenue Code

Jeffrey A. Lasky

Nova Southeastern University, [Jeffrey.Lasky@rit.edu](mailto:Jeffrey.Lasky@rit.edu)

Follow this and additional works at: [https://nsuworks.nova.edu/gscis\\_etd](https://nsuworks.nova.edu/gscis_etd)



Part of the [Computer Sciences Commons](#)

## Share Feedback About This Item

---

### NSUWorks Citation

Jeffrey A. Lasky. 2019. *The Impact of Cross-References on the Readability of the U.S. Internal Revenue Code*. Doctoral dissertation. Nova Southeastern University. Retrieved from NSUWorks, College of Computing and Engineering. (1101)  
[https://nsuworks.nova.edu/gscis\\_etd/1101](https://nsuworks.nova.edu/gscis_etd/1101).

This Dissertation is brought to you by the College of Computing and Engineering at NSUWorks. It has been accepted for inclusion in CCE Theses and Dissertations by an authorized administrator of NSUWorks. For more information, please contact [nsuworks@nova.edu](mailto:nsuworks@nova.edu).

The Impact of Cross-References on the Readability  
of the U.S. Internal Revenue Code

by


Jeffrey A. Lasky

A dissertation submitted in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy  
in  
Information Science

College of Computing and Engineering  
Nova Southeastern University

December, 2019

We hereby certify that this dissertation, submitted by Jeffrey Lasky conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirements for the degree of Doctor of Philosophy.

  
Arion B. Seagull, Ph.D.  
Chairperson of Dissertation Committee

12/5/19  
Date


  
Michael Laszlo, Ph.D.  
Dissertation Committee Member

12/5/19  
Date

  
Steven R. Terrell, Ph.D.  
Dissertation Committee Member

12/5/19  
Date

Approved:

  
Meline Kevorkian, Ed.D.  
Interim Dean, College of Computing and Engineering

12/5/19  
Date

College of Computing and Engineering  
Nova Southeastern University

An Abstract of a Dissertation Submitted to Nova Southeastern University  
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

## The Impact of Cross-References on the Readability of the U.S. Internal Revenue Code

by  
Jeffrey A. Lasky  
December 2019

Scholars and practitioners have long argued that U.S. income tax law (“the Tax Code”) is excessively complex and difficult to understand, and hence imposes non-trivial adjudication, administration, planning, and compliance costs across the spectrum of income tax stakeholders: the courts, the Internal Revenue Service, tax practitioners, business managers, and individual taxpayers. Hence, there is considerable interest in reducing the effort needed to accurately understand and apply the provisions of income tax law. Prior scholarly work has strongly argued that exceptions to Tax Code provisions as expressed by cross-references embedded in the Tax Code text constitute a major source of reading complexity.

The goal of the study was to gain a first empirical understanding about the readability impacts on users who encounter cross-references while reading Tax Code provisions. The study included a human subjects task performance experiment with 75 undergraduate and graduate accounting student participants who were completing or had completed an introductory level course in federal income taxation. Participants were presented with integrated tax scenarios and accompanying sets of scenario questions. Copies of several Tax Code sections were the only reference materials available to the study participants. The study was based on a within-subjects experimental design.

To investigate the prior work argument, cross-references embedded in the Tax Code reference materials provided to study participants that expressed exceptions were all assigned to one cross-reference category, and all other cross-references that served different purposes were assigned to a second category. As responses to scenario questions were binary (correct/incorrect), logistic regression was used to test study hypotheses.

The study’s major finding was that reading cross-references assigned to the exceptions category had a very strong negative effect on task performance, while reading cross-references assigned to the second category had a modest positive effect on task performance. The finding thus supports decades of analysis and argument that cross-references related to expressing exceptions are a major source of Tax Code reading complexity. This outcome warrants further research into statutory exception language, that subset of statutory language used to express exceptions. Such a subset will include cross-references as one of many language elements that are available for the purpose of expressing exceptions.

## Acknowledgements

I wish to extend a most sincere thank you to the members of my dissertation committee, Dr. Amon Seagull (Advisor), Dr. Michael Laszlo, and Dr. Steven Terrell, each of whom provided guidance and wisdom in their own unique manner. For those who know me best, all would be quick to agree that Dr. Seagull, with unwavering and sustained commitment to my success, pulled off a mission impossible.

I am grateful to my colleagues and friends who provided needed encouragement during the inevitable dark times.

The heroes of the story are my family, Sandi, Barley, and Violet, who endured with grace, over a prolonged period of time.

## Table of Contents

<b>Abstract</b>	iii
<b>Acknowledgements</b>	iv
<b>Table of Contents</b>	v
<b>List of Tables</b>	vii
<b>List of Figures</b>	viii

### Chapters

<b>1. Introduction</b>	<b>1</b>
Background	1
Problem Statement	24
Dissertation Goal	25
Research Question	25
Relevance and Significance	25
Barriers and Issues	27
Assumptions, Limitations, and Delimitations	27
Definitions of Terms	29
List of Acronyms	29
Summary	29
<b>2. Review of the Literature</b>	<b>31</b>
Overview	31
The Nature of Legislative Expression for Income Tax Law	33
Information Structures for Legislative Provisions	38
Linguistic Analysis of Legislative Language	42
Linguistic Analysis of Legislation	44
Network Analysis of Legislation	52
Related Human Subjects Experimentation	57
Summary and Conclusions	64
<b>3. Methodology</b>	<b>67</b>
Motivation	67
Overview of the Approach	67
Development of the Experimental Scenario	68
Experimental Design	71
Experimental Variables	73
Hypotheses	81
Data Collection and Data Preparation	82
Data Analysis	83
Study Development Activities	85

- 4. Results 89**
  - Conduct of the Experiment 89
  - Study Dataset Development 91
  - Experimental Results 93
  - Hypothesis Testing 97
  - Post-test participant experience survey 110
  - Summary and Conclusion 114
  
- 5. Conclusions, Implications, Recommends and Summary 116**
  - Introduction 116
  - Study hypotheses and conclusions 117
  - Implications 124
  - Recommendations for future work 128
  - Summary 133

## **Appendices**

- A. Participant Scenario Materials 135
- B. Scenario Tax Code Sections 154
- C. Provisions per Question and Data Tuples 168
- D. IRB Approvals 175
- E. Scenario Solutions 179
- F. Tax Expert Reviewers 188
- G. Distributions of Independent Variables 189
- H. Participant Survey 192

## **References 195**

## List of Tables

### Tables

1. Sample of Flesch Reading Ease Scores 5
2. Section 32 External Cross-References 12
3. Comparative Complexity Ranking of Title 42 and Title 26 55
4. Top Five U.S.C. Titles with Largest Cores 56
5. Composition of the Study Dataset 93
6. Undergraduate Participant Question Responses 94
7. Graduate Participant Question Responses 95
8. H1 model results 101
9. H2 model results 102
10. H3 model results 102
11. H4 model results 104
12. H5 model results 106
13. H6 model results 109
- F1. Provisions per Question 169
- F2. Question Data Tuples 173



## List of Figures

### Figures

1. Section 32, Earned Income Tax Credit, network view 12
2. IRS Commissioner scenario network view 13
3. Visualization of U.S.C. Titles with largest cores 56
4. Ranked percent correct answers 96
5. Distribution of Scope.Tokens values 99
6. Distribution of Tot\_CR.Nodes values 99
7. Study experience survey – undergraduate students 111
8. Study experience survey – graduate students 112

## Chapter 1

### Introduction

#### Background

Communities of practice shape written language to meet their communication needs and communication customs. Such shaping typically includes, for example, the development and use of specialized or technical vocabularies, the use of unusual syntactical constructions, the use of certain words and phrases as connectives and conjunctions to add cohesion to the text, and the physical organization and presentation of the text. The result of this shaping creates new language varieties, or synonymously sublanguages (Kittredge & Lehrberger, 1982). Examples of sublanguages include legal (Charrow, Crandall, & Charrow, 1982), medical and scientific, religious, and newspaper reporting (Crystal & Davy, 1969), and military (Kittredge, 1982).

In some cases, a sublanguage will be naturally subdivided into a set of more specialized sublanguages. The legal sublanguage, for example, includes a legal document language for drafting legislation (hereafter “the statutory language” or “the legislative language”) and one for drafting contracts (Trosberg, 1997). Characteristics of the statutory language include common and uncommon terms with meanings specific to legal use, long sentences, syntactic complexity, complex conditionals, chain-like structures of conditions and of exceptions to rules, and high incidences of complex prepositional phrases (Charrow, Crandall, & Charrow, 1982; Danet, 1985; Williams, 2007). Such statutory language characteristics are due, in part, to the presence of frequently occurring, complex, detailed contingencies in statutes (Bhatia, 1987, p. 9), and the requirement for

statutes to be expressed in a way that avoids ambiguity. As Trosberg (1997) remarked “A statute is a highly serious social document of ultimate authority which needs careful expression” (p. 30).

### *Codification*

During the period 1795-1926, the U.S. government wrestled with the question of how best to organize, maintain, and publish federal statutes (Tress, 2010). In 1874, the first codification of federal statutes was published as the Revised Statutes of the United States. Thereafter and over time, the statutes were re-codified and published as the United States Code (U.S.C), the present-day codification of the permanent laws of the United States. In the context of legislation, codification is the formal process whereby statutes are placed in a topically organized collection. Each statute is expressed in a hierarchical structure, where the nodes of the hierarchy contain the statute’s sections, provisions, and sub-provisions. The nodes are identified using a consistent numeric or alphanumeric labeling scheme.

### *Referential constructions*

Stevenson (2014) notes that once numeric identifiers became available, cross-referencing among and between statutes was greatly facilitated. In this context, a cross-reference is a unique linguistic component of the statutory language that establishes a legal relationship between two or more statutory provisions. Cross-references are embedded in the provisions themselves, identify another labeled provision, and are integral to expression of the provisions. Cross-reference examples include as defined in, under the provisions of, and except as provided in. Such referential constructions are

essential since “legal provisions take effect in the context of a number of other provisions and subprovisions” (Bhatia, 1983, p. 212).

The movement to codify statutes and regulations created, as an unintended consequence, a substantial increase in the use of cross-references, and arguably a corresponding increase in complexity. According to Bhatia (1983), the complex forms of syntactic constructions and referential structures found in legislative text reflect how the draftsman has uniquely used the statutory language to meet the expectations of precision and absence of ambiguity. Bhatia goes on to say that such careful and detailed expression is not without the cost “of producing a [statutory] document which may be regarded by the reader as pompous, tedious, and unnecessarily complex, and hence unreadable” (p. 24).

Although Stevenson (2014) did not explicitly say so, the decision to codify was an information design decision. The study is centrally concerned with examining claims about unintended consequences of this information design decision on statutory readability and usability, with particular focus on cross-references as one element of statutory expression and presentation.

In the following section, the matters of readability and usability of complex statutory text are examined with specific reference to the Internal Revenue Code, the statute that expresses U.S. income tax law. The Internal Revenue Code is widely considered to be the leading example of statutory complexity, and has, for decades, received scholarly attention, and less rigorous attention as well, from stakeholders in the public and private sectors.

*Captivation with the complexity of income tax law*

Considerations of income tax law are often central in the planning and decision-making activities of individuals, families, and organized entities of all types. Hence, income tax law affects large numbers of people and organizations. Yet, income tax law in the United States, as well as in other English-speaking countries such as Australia, Canada, New Zealand, and the United Kingdom, has long been, and continues to be, assailed as being numbingly complex. For the purposes of this study, complex means challenging and difficult to read, to understand, and to apply. Consideration of other forms of income tax complexity, including income tax policy, compliance complexity, and administrative complexity, are outside the scope of the study.

Legal writings, as typically expressed in legalese, are considered to be among the more complex types of documents (Benson, 1985). Some supporting evidence for this assertion is found in Flesch (1979a). Flesch, the developer of the nearly ubiquitous Flesch Reading Ease score, randomly sampled writings from a variety of sources, and then calculated the Reading Ease score for each. The results are shown in Table 1 below (some omitted, bold emphasis added to denote legal discourse) (p.26). The score is calibrated to a scale of 1 to 100, where a value of 100 categorizes text that is easy to read, and a value of zero identifies text that is practically unreadable. The Internal Revenue Code is ranked essential unreadable.

Table 1.

## Sample of Flesch Reading Ease Scores

Sample publication	Average Reading Ease Score
Comics	92
Consumer ads in magazines	82
Movie Screen	75
Readers Digest	65
Time	52
Harvard Business Review Wall Street Journal	43
The New York Review of Books	35
<b>Harvard Law Review</b>	<b>32</b>
<b>Standard auto insurance policy</b>	<b>10</b>
<b>Internal Revenue Code</b>	<b>-6</b>

Commentaries on income tax legislation frequently invoke a sense of despair and frustration related to difficulties in understanding income tax law as expressed in the form of statutory writings. The following comment, attributed to a judge's remarks written in 1990 about two subsections of the Australian Income Tax Assessment Act of 1936, is a representative example:

While both subsections present difficulties of construction, the former is drafted [written] with such obscurity that even those used to interpreting the utterances of the Delphic oracle might falter in seeking to elicit a sensible meaning from its terms (cited in Williams, 2007, p.14)

Other types of comments sometimes provide indications as to possible sources of such understanding and application difficulties. The now classic remarks of Learned Hand (1947), a well-respected judge and judicial philosopher, delivered in the context of a eulogy for a colleague, provide reflective insights into the burdens associated with reading and understanding even what was then a much smaller-scale, 137-page tax law (Revenue Act of 1938) compared to the present day, approximately 4,000-page U.S. Internal Revenue Code (emphases and line spacing added).

In my own case the words of such an act as the Income Tax, for example, merely dance before my eyes in a meaningless procession: *cross reference to cross-reference, exception upon exception* - couched in abstract terms that offer no handle to seize hold of - leave in my mind only a confused sense of some vitally important, but successfully concealed, purport, which it is my duty to extract, but which is within my power, if at all, only after the most inordinate expenditure of time. ...

Much of the law is now as difficult to fathom, and more and more of it is likely to be so; for there is little doubt that we are entering a period of increasingly detailed regulation, and it will be the duty of judges to thread the path - for path there is - through these fantastic labyrinths. ... (p. 169).

In the view of Miller (1993) and others (Potter, 2004), Hand's labyrinths would also be characterized as a Serbonian Bog, where escape from entangling circumstances is problematical at best. Reference to this mythical swamp in ancient northern Egypt has a substantial history of use in legal documents (Potter, 2004).

Similar to Judge Hand's mention of cross-references and exceptions as contributors to [his] income tax law readability difficulties, Rook (1993) selected exceptions, cross-references, (and verbal descriptions of arithmetic expressions), as a framework for organizing his drafting suggestions to improve the readability of complex legislation in general, and of the Internal Revenue Code in particular. He noted that nested exceptions and chains of cross-references were especially problematical. Rook ends his introductory remarks with tongue-in-check expression of cautious optimism:

Thus, I assume that even though the statute is unartfully drafted, and therefore more complex than it need be, a reader can divine the legislation's meaning after some analysis and perhaps some aspirin (p. 666).

Less colorful complexity analysis can be found in the recent trend to apply network analysis to statutes. In the case of U.S. statutes, while the analysis is typically conducted on groups of statutes, the Internal Revenue Code is considered as a group of one. Katz and Bommarito (2014), in apparently the first work to develop a statutory

complexity measure based on network analysis, ranked the Internal Revenue Code as the second most complex federal statute. A review of their work presented in Chapter II arguably ranks the Internal Revenue Code as the most complex statute. The results of another network analysis also point to an Internal Revenue Code ranking of most complex (Li, Faz, Larochelle, Hill & Lo, 2015).

*The complex technical structure of the Internal Revenue Code*

Given the nearly continual captivation among many about income tax complexity, McCaffery (1990) asks “But what then explains our [U.S.] tax system’s persisting complexity?” (p. 1267, Abstract). Stanley Surrey, the U.S. Assistant Secretary of the Treasury for Tax Policy from 1961-1969, and a professor at Harvard Law School, offers characterizations of U.S. income tax law that can serve as an organizing framework for the explanations that McCaffery seeks.

... a complex technical structure ... [comprised of] complex substantive tax rules with complex interrelationships, characterized by complex variations in the tax treatment of transactions often not differing greatly in substance or form, all of which are expressed in a complex statutory terminology and arrangement (Surrey, 1969, p.673).

Surrey continues with describing the relationships between a reader and the complex technical structure:

The sheer complexity of the income tax, with its countless artificial rules and definitions, involves so many possible combinations of rules that efforts to study and evaluate the possible solutions to any single problem become enormously difficult. As a problem emerges, it appears to have a certain contour. ... Soon further study begins to disclose that the problem does not stand in isolation. It is connected by many strands with other difficulties. It is but a surface manifestation of deeper troubles. Its conceptual significance then becomes clouded as the links with other concepts begin to emerge (p. 687).

Taken together, Surrey’s two descriptions reveal two high-level, interrelated properties of the complex technical structure, a static dimension and a dynamic



dimension. In the context of reading and understanding, McCaffery (1990) defines the term technical complexity as “the pure intellectual difficulty of ascertaining the meaning of tax law ... a relatively static matter concerning the understandability of a particular Tax Code section in abstract isolation from other sections” (p. 1271). The term Tax Code section identifies the entire collection of income tax statutory text that is physically assembled, presented, and identified with a section number and title.

The dynamic dimension is associated with applying the information encapsulated in the complex technical structure to tax problem solving or to tax planning. For these purposes, many sections of the Tax Code likely will be consulted, and perhaps sections of other statutes as well. Arguably, these tasks present higher levels of intellectual difficulty compared to the static case. McCaffery continues, “In contrast, structural complexity is dynamic in nature in that it involves the transactional effects of a tax rule in the context of the Code as a whole” (p. 1271).

A parse of Surrey’s characterizations serves to identify properties of the static and dynamic dimensions. Static properties include arrangement, rules and expression of rule interrelatedness, and language. Dynamic properties include problem definition and dimensionality, scale of the rules space, concept and solution emergence, discovery of related concepts, and solution elaboration. The following sections examine several of the properties that are associated with the static and dynamic dimensions of the complex technical structure, and their potential contributions to Tax Code complexity.

There are two discernable information structures, one hierarchical and the other network, that form the foundation of the complex technical infrastructure. The hierarchical information structure contains a table of contents and the text of the statute’s

provisions. The network information structure is comprised of a (large) set of cross-references that express relationships between the statutory provisions.

### *Hierarchical structure*

The text of the general and permanent U.S. federal statutes is compiled and codified in the United States Code (U.S.C.). The U.S.C. is hierarchically structured, where the top level organizes federal statutes into Titles that represent broad areas of law. In some cases, such as Title 42 - The Public Health and Welfare, numerous, disparate statutes are organized under one Title. In other cases, such as Title 26 - Internal Revenue Code (the "Tax Code"), the Title contains only one statute. Presently (as of 2018) 53 Titles comprise the U.S.C.

Each statute is hierarchically organized as a topic hierarchy followed by a content hierarchy. The topic hierarchy is comprised of varying numbers of subtitles, chapters, subchapters, parts and subparts. For example, Title 26 includes Subtitle A (Income Tax), Subtitle B (Estate and Gift Taxes), and nine additional subtitles. Subtitle A includes Chapter 1, Normal Taxes and Surtaxes, Chapter 2, Tax on Self-Employment Income, and five other Chapters. Chapter 1 includes Subchapter A (Determination of Tax Liability), Subchapter B, (Computation of Taxable Income), and 23 additional Chapters. Each chapter may in turn be subdivided into subchapters, parts, and subparts. Hence, the Title 26 topic hierarchy is a collection of hierarchies, each beginning at Title 26 and ending at a subtitle, chapter, subchapter, part, or subpart node.

The terminal node of each topic hierarchy is followed by an associated content hierarchy. The content hierarchy contains the actual statutory text organized into varying numbers of sections, subsections, paragraphs, subparagraphs, clauses, and subclauses.

Each subdivision of the content hierarchy is labeled with a standardized U.S.C. string of upper and lower case alphanumeric characters that encodes the hierarchical structure used to organize the legislative text. The complete path to a statutory provision begins at a section encoded as a unique integer, followed as needed by a lowercase alpha character (subsection), an integer (paragraph), an uppercase alpha character (subparagraph), a lower case Roman numeral (clause), and an upper case Roman numeral (subclause), each enclosed in parentheses. For example, the complete path to section 401, subsection (a), paragraph (5), subparagraph (D), clause (i), subclause (I) will be in the form section 401(a)(5)(D)(i)(I).

#### *Network structure*

Statutes in the U.S.C. contain referential structures, or cross-references, whose purpose is to explicitly specify relationships between the statutory provisions. These cross-references are motivated by the presence of topic relatedness, or overlap, in the statute. The cross-references are expressed using the alphanumeric path format described above. The aggregate of these cross-references creates a graph, or network information organization, that is integrated into and superimposed on and across the hierarchical information organization. In graph terminology, each element of the content hierarchy (section, subsection, paragraph, subparagraph, clause, subclause) is a node, and each cross-reference is an edge that connects two nodes.

The degree of overlap, even for seemingly ordinary tax matters, can be surprisingly considerable. For example, Luttman, Monarchi, and Nagy (2015) conducted a network analysis of Tax Code section 32, the earned income credit. Section 32 contains cross-references to 12 other Tax Code sections. Figure 1 displays the network created by

the cross-references (Luttman et al., 2015, p. 23). Shown are the edges connecting section 32 to the 12 other sections, as well as additional edges that signify relationships among the 12 other sections themselves. The arrowheads in Figure 1 show the **from-to** direction of the cross-references from section to section. More generally, both the **from-provision** and the **to-provision** locations can be at any position in the provision hierarchy {section, subsection, paragraph, subparagraph, clause, subclause}.

Table 2 identifies the topic name of each section included in the Figure 1 graph. The topics represent a wide range of tax matters. For any practical set of circumstances that bear on the section 32 tax credit, only a subset(s) of the graph and the associated topics will need to be consulted. There are, however, many sets of potential circumstances that will lead to consideration of other graph subsets and their associated topics. The graph depicts only the inter-section, or external cross-references, those that formally define relationships between the provisions of two or more different sections. Not displayed are the intra-section, or internal cross-references, those that formally define relationships among the provisions of one section.

There may also be extra-statute cross-references, those that formally define relationships between different statutes. From the network perspective, one of the distinguishing properties of the Internal Revenue Code is that 97 percent of cross-references are internal cross-references (Katz & Bommarito, 2014). Thus, the Internal Revenue Code is highly self-referential, the complexity implications of which are discussed in Chapter 2.

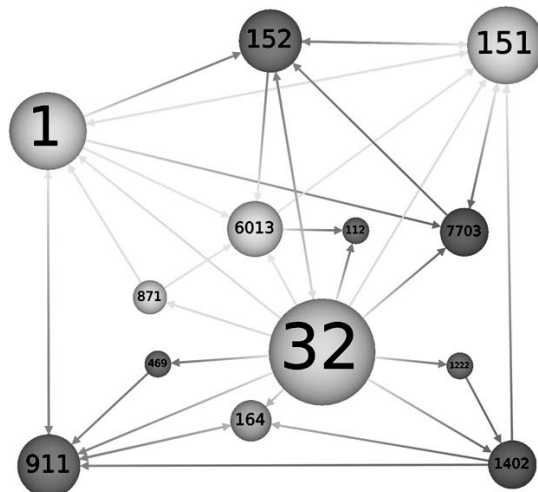


Figure 1. Section 32, Earned Income Tax Credit, network view. The relative size of circles is proportional to the number of incoming and outgoing cross-references. (Luttman, et al., 2005, p.23).

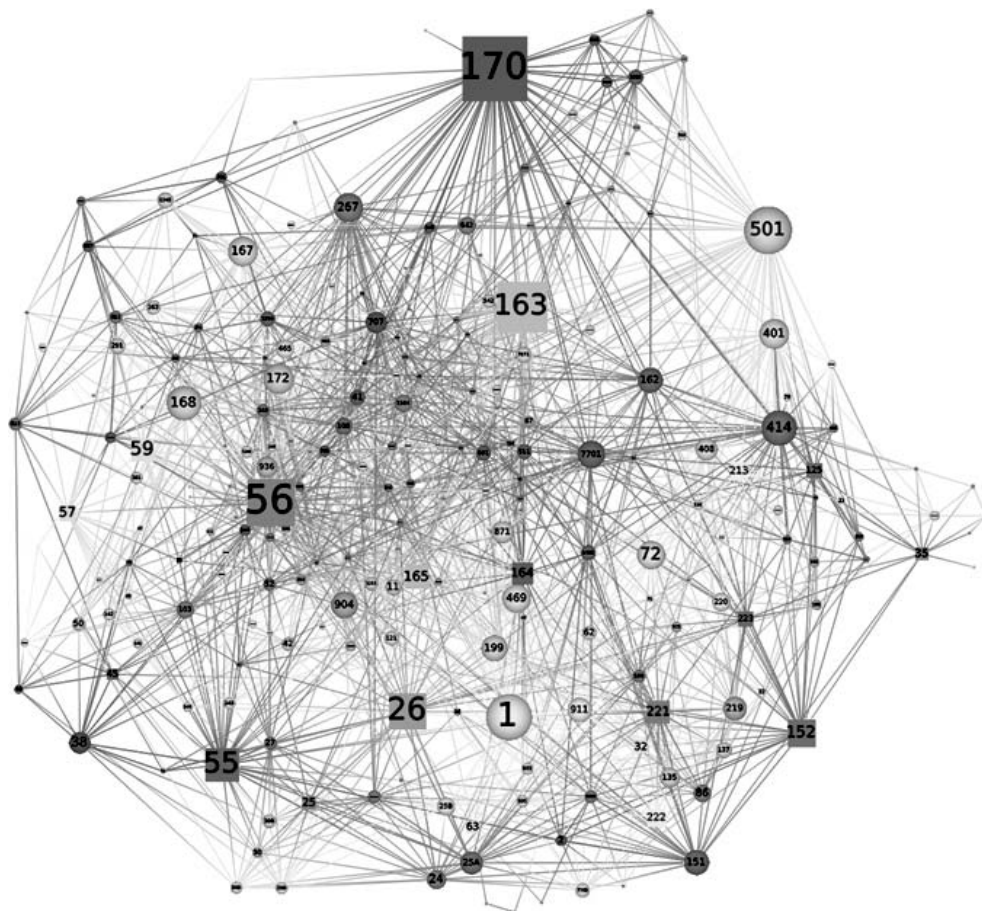
Table 2.

Section 32, External Cross-References

Section	Topic	Section	Topic
1	Taxes imposed	871	Non-resident alien individuals
112	Combat zone	911	U.S. citizens or residents living
151	Personal exemptions	1222	Capital gains and losses
152	Dependent defined	1402	Self-employment income
164	Itemized deductions	6013	Joint returns
469	Limitations on passive	7703	Determination of marital status

A more robust view into the network component of the complex technical structure, and one that illustrates the potential scale of the technical structure’s complexity, is provided by a network analysis of a tax scenario presented by a former IRS Commissioner while delivering a speech about the complexity of the Tax Code (Luttman, et al., 2015). On the surface, the scenario environment was simple: a married couple, both working, with children. The Commissioner’s modest scenario elaborations were directed mostly to common family tax matters, with some special emphasis on tax

benefits related to children. The researchers mapped the Commissioner's example scenario to relevant sections of the Tax Code. This activity identified 27 sections. The analysis was then extended to 229 other sections that referred to one or more of the 27 sections. The resulting graph is shown in Figure 2. Again, only the external cross-references are depicted in the graph.



*Figure 2.* The initial 27 sections are depicted as squares, and the additional 229 sections as circles. There is a total of 1,858 incoming and outgoing cross-references between the 256 sections. The total number of cross-references is conservative, since multiple references between two sections were counted as one intersection reference, and all intra-section references were ignored. (Luttman et al., 2015, p. 28). Shadings are the grey scale renditions of the color figure as published.

*Structure interaction*

Topic interrelatedness, and the cross-references that formally create the interrelationships, are thought to be a source of reading and comprehension impairment (Burton and Karlinsky, 2016, p. 248). As a reader naturally traverses the statutory text hierarchy in linear and downward direction, the reader can expect that the next node will provide detail related to the information in the immediately above node, as this is a fundamental property of an information hierarchy. In other words, there will be no information surprises that could disrupt a reader's train of thought. (This is not to say, however, that hierarchical information structures will necessarily spare a reader from comprehension challenges. A reader's train of thought can also be disrupted by exposure to successive nodes of increasing detail, absence of surprise notwithstanding).

In a pure hierarchy, there are no linkages that connect the discrete hierarchical paths. However, the cross-references embedded in the statutory text do serve as connections between discrete hierarchical paths. Hence, statutes codified in the U.S.C are not organized as pure hierarchies, but rather are organized as composite structures, part hierarchical and part network.

The composite information structure does not have the information topic predictability of a pure hierarchy, and so may contribute to comprehension difficulties. This is the case because traversing the path formed by an external cross-reference(s) will move the reader's attention to another discrete branch(s) in the hierarchy, and hence to another topic or concept. If the path is formed by an internal cross-reference, the reader's attention will also be impacted. These attention events may cause the reader to experience a coherence disruption that will impact the reader's thought processes and thus degrade

understanding (Miller & Kintsch, 1980). In addition, and importantly, different from the serial, uninterrupted reading mode associated with reading a hierarchical structure, a reader of the composite structure will experience an interruption each time a cross-reference is encountered in the text. Interrupted reading has been shown to cause comprehension difficulties (Foroughi, Werner, Barragan, & Bohm-Davis, 2015).

*Expression of rules and rule interrelationships*

Statutes express the law in terms of primary rules and related secondary rules. The primary rules express a provision of law in relatively general terms. The secondary rules specify the conditions under which the general rule will and will not be operative, and if operative, whether the terms of the general rule will be adjusted due to circumstances. Hence, the expression of a specific point of law is a union of primary and secondary rules (Hart, 1994), or using Bhatia's terminology (1983), the union of main provisions and qualification provisions.

The secondary rules, or qualification provisions, provide definitions, and also set forth exceptions, limitations, special rules, and ordering instructions for the application of provisions. The qualification provisions are expressed in one of two ways. In the first, the qualifications are inserted into the statutory text as self-contained provisions; i.e., the text does not include a cross-reference(s). In the second, the qualification provisions do include a cross-reference(s). The referred to provision(s) may also include a cross-reference(s), thereby creating a chain(s) of cross-references.

The legislative draftsman is responsible for transforming the intent of the legislature into written statutory law. The draftsman relies on cross-references to accomplish two major goals:



1. To reduce the volume of the legislative text by referencing an already existing definition rather than repeating the definition in each provision where the defined term is mentioned.

2. To explicitly and formally record legal relationships among and between provisions, not only at the time the statute is initially drafted, but as well over time as the statutory text is amended thereby potentially requiring cross-reference additions, modifications, and deletions. McCaffery (1990) notes one primary motivation for changes to tax legislation is “taxpayer exploitation of gaps, ambiguities, and inconsistencies in the rules” (p. 1277) as taxpayers seek to minimize their tax liability. Flesch (1979b) provides a detailed mini-history of how cross-reference use in the Tax Code evolved in reaction to taxpayer attempts to minimize investment tax liability by taking advantage of how the Tax Code defined family members.

The cross-references that explicitly create and signal interrelationships between the statute’s provisions are expressed in a unique syntactical style. A cross-reference is a composite syntactical unit comprised of a natural language cue word or phrase together with a pointer to a location the statutory text. The cue portion of the cross-reference, italicized in the following examples, gives a reader a context or a processing instruction for the information in the referenced text. For example, *as defined in* section 63, and *determined without regard to* section 152(e), where in the former, the reader is provided with the location of a definition, and in the latter, the reader is being instructed to disregard the rule(s) contained in section 152(e). The legislative draftsman in effect types the graph edge by supplying semantic information in the form of cue phrases.

Edge, or cross-reference, typing is fundamental to the expression of statutory text, since the conditions under which a particular provision will apply, or will not apply, are variable and sometimes nuanced. The draftsman captures this state of affairs, for example, by authoring cue phrases such as *does not apply by reason of*; *shall be applied as if*; *shall apply in lieu of the rules of*; and *shall apply only as provided by*.

### *Complex statutory terminology*

Here, the interpretation of complex statutory terminology as a component of Surrey's description of the Internal Revenue Code as a complex technical structure, is twofold. The first component, mentioned previously, is the presence of unusual syntactic constructions present in the statutory text. Benson (1985) presents an illustrated list of such constructions labeled as *syntactic features* (p. 524). Bhatia (1983) includes an extended discussion of what he termed syntactic discontinuity, a problematic consequence of a particular and common drafting style that attempts to place qualification clauses immediately following the word or phrase that is the qualification target. The normal processing of the sentence is then disrupted (p. 229).

The second component is the often-encountered complexity of how certain terms are defined. Morrison (2014) notes that the inclusion of defined terms, as well as cross-references, are essentially inevitable in a rules-driven, complex statute such as the Internal Revenue Code. The need to include defined terms is a direct result of using specialized terminology to express the rules that comprise legislative provisions. For certain terms, the volume of detail a reader must absorb can be overwhelming, as the recursive-like processing of nested terms and nested cross-references will likely prove burdensome to a reader. In Chapter 2, Morrison's central example is presented and

extended to illustrate the degree of complexity that is attached to how a term such as *qualified retirement plan* is defined.

*The nature of reading the Tax Code*

Tax Code §117, Qualified Scholarships, serves to illustrate the nature of reading the tax code. Here are presented, in tutorial form, the principal attributes of the statutory writing style. The following reproduced section 117 is based on a file prepared by the Office of Law Revision Council, U.S. House of Representatives. This office is responsible for maintaining the United States Code. The discussion approach is to review each section 117 provision in the context of the material presented up to this point.

## **§117. Qualified scholarships**

### **(a) General rule**

Gross income does not include any amount received as a qualified scholarship by an individual who is a candidate for a degree at an educational organization described in section 170(b)(1)(A)(ii).

### **(b) Qualified scholarship**

For purposes of this section-

#### **(1) In general**

The term "qualified scholarship" means any amount received by an individual as a scholarship or fellowship grant to the extent the individual establishes that, in accordance with the conditions of the grant, such amount was used for qualified tuition and related expenses.

#### **(2) Qualified tuition and related expenses**

For purposes of paragraph (1), the term "qualified tuition and related expenses" means-

(A) tuition and fees required for the enrollment or attendance of a student at an educational organization described in section 170(b)(1)(A)(ii), and

(B) fees, books, supplies, and equipment required for courses of instruction at such an educational organization.

### **(c) Limitation**

#### **(1) In general**

Except as provided in paragraph (2), subsections (a) and (d) shall not apply to that portion of any amount received which represents payment for teaching, research, or other services by the student required as a condition for receiving the qualified scholarship or qualified tuition reduction.

**(2) Exceptions**

Paragraph (1) shall not apply to any amount received by an individual under-

(A) the National Health Service Corps Scholarship Program under section 338A(g)(1)(A) of the Public Health Service Act,

(B) the Armed Forces Health Professions Scholarship and Financial Assistance program under subchapter I of chapter 105 of title 10, United States Code, or

(C) a comprehensive student work-learning-service program (as defined in section 448(e) of the Higher Education Act of 1965) operated by a work college (as defined in such section).

**(d) Qualified tuition reduction****(1) In general**

Gross income shall not include any qualified tuition reduction.

**(2) Qualified tuition reduction**

For purposes of this subsection, the term "qualified tuition reduction" means the amount of any reduction in tuition provided to an employee of an organization described in section 170(b)(1)(A)(ii) for the education (below the graduate level) at such organization (or another organization described in section 170(b)(1)(A)(ii)) of-

(A) such employee, or

(B) any person treated as an employee (or whose use is treated as an employee use) under the rules of section 132(h).

**(3) Reduction must not discriminate in favor of highly compensated, etc.**

Paragraph (1) shall apply with respect to any qualified tuition reduction provided with respect to any highly compensated employee only if such reduction is available on substantially the same terms to each member of a group of employees which is defined under a reasonable classification set up by the employer which does not discriminate in favor of highly compensated employees (within the meaning of section 414(q)). For purposes of this paragraph, the term "highly compensated employee" has the meaning given such term by section 414(q).

**[(4) Repealed. Pub. L. 101-140, title II, §203(a)(1), (2), Nov. 8, 1989, 103 Stat. 830 ]**

**(5) Special rules for teaching and research assistants**

In the case of the education of an individual who is a graduate student at an educational organization described in section 170(b)(1)(A)(ii) and who is engaged in teaching or research activities for such organization, paragraph (2) shall be applied as if it did not contain the phrase "(below the graduate level)"

*§117 Narrative and Commentary*

Typically, a Tax Code section will begin with a general rule. Subsection (a) is the general rule and authority that excludes from gross income amounts received from a qualified scholarship by a degree candidate at an educational organization that meets the criteria in section 170(b)(1)(A)(ii). *Qualified scholarship* and *educational institution* are examples of the large number of defined terms in the Tax Code. The definition of a term

will generally be available in the same section where the term occurs, or in a different section in the Tax Code. In some cases, the definition will be available in a different statute or law, or in another legal writing external to the Tax Code. Here, qualified scholarship is defined in subsection (b), while the definition of educational institution is obtained from the cross-reference to section 170(b)(1)(A)(ii). The section 170(b)(1)(A)(ii) cross-reference is an example of an external cross-reference.

Subsection (b)(1) contains the definition of qualified scholarship and introduces the defined term *qualified tuition and related expenses*. Subsection (b)(1) also imposes the requirement that for income exclusion, the scholarship amounts must actually be used for qualified tuition and related expenses. (To the extent that the requirement is not satisfied, such amounts are included in gross income).

Subsection (b)(2) contains the definition of qualified tuition and related expenses. Note the phrase *For purposes of paragraph (1)* at the beginning of subsection (b)(2). The phrase *For purposes of* serves to limit the scope of the definition of qualified tuition and related expenses to the context of subsection (b)(1). The definition presents two categories of qualified expenses found in subparagraph (A) and subparagraph (B). Note a second identical use of a cross-reference to section 170(b)(1)(A)(ii). The repeated use of a cross-reference within a section is a common element of the legislative writing style, although not a principal element.

Subsection (c)(1) imposes a limitation on the general scholarship gross income exclusion rule. Scholarship amounts that represent compensation for required teaching, research, or other services are not excludable from gross income. Note the writing approach used to express the limitation. First, the excludable scholarship amount is

identified by the phrase *Except as provided* in paragraph (2). The reader will need to resolve the internal cross-reference to paragraph (2); the paragraph (2) in (b)(1)(2) or in (c)(2)? The resolution rule is that in the absence of additional path information, resolution takes place in the most immediate context, in this case, subsection (c).

Subsection (c)(2), Exceptions, again identifies excludable amounts by referencing that paragraph (1) will not apply to any amounts, including payments for services, included in certain scholarship and other education programs listed in subsection (c)(2)(A), (B), or (C). Second, the provisions that authorize the general scholarship exclusions, subsections (a) and (d), will not be applicable to any amount that represents compensation for services (except compensation for services included in subsection (c)(2)(A), (B), or (C) scholarships).

Subsection (d) introduces the defined term *qualified tuition reduction*, another form of excludable educational support under section 117. Subsection (d)(1) is the general rule and authority for excluding qualified tuition reduction amounts from gross income. Subsection (d)(1)(2) defines the term *qualified tuition reduction*. This benefit is available to an employee of a section 170(b)(1)(A)(ii) organization only for education below the graduate level (but see subsection (d)(5)). By virtue of the cross-reference to section 132(h), Certain Fringe Benefits, that is embedded in subsection (d)(2)(B), the definition of employee is expanded to include retired and disabled employees of the organization, a surviving spouse of the employee, the employee's current spouse, and the employee's dependent child or children. Section 132(h)(2)(B) contains a cross-reference to section 152(f)(1), Child defined. According to section 152(f)(1), the term *child* includes only son, daughter, stepson, stepdaughter, and a foster child.

Subsection (d)(3) is a non-discrimination rule that denies the exclusion benefit if the tuition reduction plan is only available to highly compensated employees. The method to impose the denial is a condition placed on the applicability of paragraph (1), the general exclusion from gross income rule for qualified tuition reduction. The defined term *highly compensated employee* is defined in cross-referenced section 414(q), Highly Compensated Employee. Section 414(q) includes nine paragraphs, five external cross-references, and 13 internal cross-references.

Subsection (d)(4) is repealed.

Subsection (d)(5) is a special rule for teaching and research assistants that modifies the language of paragraph (2) that includes the phrase “(below the graduate level)”. For a graduate student who is a teaching or research assistant, the rules for gross income exclusion of qualified tuition reduction will apply. The method to enact the special rule is the subsection (d)(5) language: paragraph (2) shall be applied as if it did not contain the phrase “(below the graduate level)”.

#### *§117 Narrative and Commentary Revisited*

The purpose of this section was to provide a commented experience of reading a (brief) Tax Code section, and to illustrate the principal attributes of the statutory writing style. A reading of section 117(d) provides a glimpse into the complexity of the Tax Code. Section 117(d) includes an external cross-reference (to section 132(h)) that expands the common meaning of employee, an external cross-reference to section 414(q) that in turn has external cross-references to five other sections, and an example in paragraph (d)(5) of one provision modifying the text of another provision in the same

subsection by use of a backward internal cross-reference from paragraph (d)(5) to paragraph (d)(2).

In addition, it is seen that cross-references will serve different purposes. Bhatia (1983) developed a four-category classification of cross-references purposes. Here, each category is identified and illustrated with examples from the section 117 text.

Providing textual authority. Subsections (a) and (d) provide the authority to exclude from gross income amounts received as a qualified scholarship, and qualified tuition reductions, respectively.

Referring to terminological explanation. The subsection (a) cross-reference to section 170(b)(1)(A)(ii) references the definition of educational organization; the paragraph (d)(3) cross-reference to subsection 414(g) references the definition of highly compensated employee.

Providing textual mapping. Textual mapping cross-references establish a text-cohering relationship for reducing information processing load at specific points in provisions. Typically, instances of textual mapping cross-reference are seen in paired cross-references. Paragraph (c)(1) contains a forward reference to paragraph (c)(2), while paragraph (c)(2) contains a backward reference to paragraph (c)(1). In this example, a reader can review the rule in paragraph (c)(1) while deferring a review of the specific exceptions contained in paragraph (c)(2); also, the backward pointer in paragraph (c)(2) serves as a contextual reminder of the general rule of paragraph (c)(1).

Defining legal scope. Legal scope is a determination of how one provision can or will affect the operation of another provision. For example, paragraph (d)(2) constrains the application of the paragraph (d)(1) general rule to undergraduate students. However,



the paragraph (d)(5) special rule serves to rescind the constraint for graduate students who serve as teaching or research assistants.

It is an open question whether the four different types of cross-references will have the same or different impacts on Tax Code readability.

### *Conclusion*

The issue of the degree of complexity present in the U.S. Tax Code, considered in the context of reading, comprehension, and application challenges, is long standing. Surrey's conceptualization of the U.S. Tax Code as a complex technical structure is supported to certain extents by prior work, albeit sometimes accompanied by emotional outbursts. The properties of legal languages in general, and the statutory language in specific, continue to receive academic attention. On the other hand, there has been less academic attention, from the behavioral perspective, focused on whether and to what extent the use of cross-references to express statutory provisions contributes to long-standing Tax Code complexity concerns.

### **Problem Statement**

Notwithstanding commentary from income tax scholars and practitioners about the adverse and potentially severe readability effects associated with the presence of cross-references in the U.S. Tax Code, only minimal prior scholarly work has been directed toward increasing understanding about the extents to which the presence of cross-references in income tax legislation impair reading and comprehension, and impair capabilities to apply specific Tax Code provisions to income tax problems.

More recently, however, as will be discussed in Chapter II, increasing research attention is being directed to the identification and classification of cross-references in

legislative text, to the relationship between cross-references and the complex technical infrastructure, and more generally, to readability complexity. The preponderance of this work is in the form of computer-based linguistic and network analysis applied to statutory text corpora. Parallel research efforts in the form of human-subjects experimentation for investigating the potential relationships between cross-references embedded in statutory text, and statutory text readability, can serve as a useful, complementary avenue of research.

### **Dissertation Goal**

The goal of the study was to gain a first human-subject based understanding about the extent to which the presence of cross-references in the Tax Code impairs reading, comprehension, and applications of Tax Code provisions.

### **Research Question**

To what extent do cross-references contribute to a reader's intellectual difficulty of collecting and synthesizing information within and across sections of income tax law text?

### **Relevance and Significance**

The complexity of U.S. income tax law imposes extremely large adjudication, administration, planning, and compliance costs across the spectrum of income tax stakeholders: the courts, the Internal Revenue Service, tax practitioners, business managers, corporations, non-profit organizations, and individual taxpayers. The most recent National Taxpayers Union Foundation's estimate of annual Title 26 compliance costs for individuals and corporations is \$364 billion, the majority of which, \$273 billion, is the time value of the approximately eight billion hours invested in tax related activities

(Brady, 2019). The remaining \$91 billion is out-of-pocket costs for professional services and software.

As noted earlier, the Tax Code is continually evolving. Each addition or revision to a section of the Tax Code carries the potential for new cross-reference use, and hence potential for additional technical complexity and costs. This research will develop new knowledge about the impacts of cross-references on reading and comprehending legal text. There is high potential for generalization of the results to other statutes included in the U.S.C., and perhaps to other codified collections of statutes and regulations. However, this generalization potential needs to be tempered as already completed network studies of the U.S.C. Titles show considerable variations in structural properties and topic breadth (Katz & Bommarito, 2014; Li et al. 2015). Studies similar to the study, for example, that focus on other U.S.C. statutes will need to be conducted before claims about generalization can be considered. There also is high potential for original work, since apparently there are only a few extant human subjects experimental studies that address the relationship between the presence of cross-references and their impact on statutory readability.

The long-range goal of this work is to eventually develop sufficient experimental results to support the design of a legal drafting tool. The envisioned tool would simulate and visualize the complexity effect of a proposed statutory change to inform the draftsman of the consequences of the change. Along these lines, it may be possible to develop a complexity index that would rank provisions by complexity, thereby identifying provisions that are candidates for revisions to reduce reading complexity.

## **Barriers and Issues**

The dissertation problem is directed toward a corpus of income tax statutory text that is generally viewed as complex, difficult to read, and difficult to understand. Only minimal prior work has been conducted on designing a human subjects experiment to determine the impact of cross-references on readability. Hence, development of the study took place largely without access to directly related prior work.

## **Assumptions, limitations, and delimitations**

### *Assumptions*

- Student participants will be able to successfully apply their taxation knowledge and taxation problem-solving approaches they have acquired from coursework, and in some cases from work experience as well, to the experimental task. Included in this assumption is that student participants are will abandon reading the immediate provision as soon as a determination, or decision, can be made.
- Student participants will respond in a thoughtful manner when responding to the post-test questionnaire regarding their experience with the experimental task.

### *Limitations*

- All study participants were accounting students enrolled in required courses. Educational experiences in the introductory federal taxation course that all undergraduate student participants were in the process of completing, or in the case of graduate students had completed, may differ across different colleges/universities.
- Students had the opportunity to withhold permission for inclusion of their work on the experimental task in the study dataset. Four of the 83 course enrollees withheld permission.

- Student participants may silently have abandoned the experimental task due to frustration with the difficulty of the questions, or due to fatigue.
- Student participants may have engaged in variable amounts of undetectable Tax Code provision rereading, thereby introducing uncontrolled variation in reading volumes per participant.
- Student participants may have ignored the provided reference materials when responding to experimental task questions.
- Student participants may have engaged in guessing behavior for the Y / N type experimental task questions.

#### *Delimitations*

- The experimental task was based on an arbitrary selection of Tax Code sections, and on an associated set of arbitrarily drafted tax scenario questions. Experimental results may have differed with different selections of Tax Code sections and/or with different questions.
- The experimental task was based on the federal income tax statute as codified in the United States Code. Experimental results may have differed if the experimental task had been based on other federal statutes also codified in the United States Code.
- Experimental results may have differed if the experimental task was based on the income tax statutes of other English-speaking countries.
- Introductory undergraduate courses in federal taxation typically do not expose students to the Tax Code statutory language. Students enrolled in M.S. in Taxation programs and students enrolled in law school will have substantial exposure to statutory language. Experimental results may have differed with different student populations.

- Individuals with professional taxation domain expertise, such as Certified Public Accountants and Tax Attorneys, would likely had exhibited different task performance levels.

### **Definition of terms**

*Title 26 – Internal Revenue Code*, is the formal identifier of the U.S. statute that includes, among other topics, income taxes, estate and gift taxes, employment taxes, alcohol, tobacco, and other excise taxes, and the financing of presidential election campaigns.

*United States Code* “is a consolidation and codification by subject matter of the general and permanent laws of the United States. It is prepared [and maintained] by the Office of the Law Revision Counsel of the United States House of Representatives”

([uscode.house.gov](http://uscode.house.gov)).

### **List of acronyms and symbols**

IRC	Internal Revenue Code
IRS	Internal Revenue Service
NLP	Natural Language Processing
NSU	Nova Southeastern University
U.S.C.	United States Code
§	Section

### **Summary**

This chapter provided background material that showed widespread and longitudinal concerns, sometimes expressed with emotional intensity, about the readability of statutory writings in general, and of Title 26, the U.S. Tax Code in particular. The requirements to express statutory provisions with precision and with

absence of ambiguity have motivated draftspersons to adopt unique writing practices that apparently do not consider readability objectives. One of these practices, the use of cross-references to express legal relationships among and between statutory provisions, is considered by many to be a source of statutory and readability complexity. These cross-references interact with hierarchical organizations of provisions that result in the creation of graph structures. Recent studies of these graph structures reveal a wide range of structural and linguistic complexities. An analysis of one U.S. Tax Code section, §117, was presented to illustrate, albeit at small scale, how cross-references are used to express statutory provisions. Two graph visualizations illustrated how structural complexity created by cross-references scales rapidly from an example with several Tax Code sections to a larger example with 256 Tax Code sections. The question that the study investigated was to determine the impact of these linguistic and structural complexities on Tax Code readability. The next chapter reviews a broad range of prior work that is related to this question.

## Chapter 2

### Review of the Literature

#### Overview

Chapter 1 presented rhetoric and more reasoned concern about the complexities associated with reading, understanding, and applying income tax law. Three categories of potential sources of complexity were identified: the structural organization of legislative writings, attributes of the legislative language, and the pervasive use of embedded cross-references to express relationships between legislative provisions. Surrey (1969) used the term “complex technical structure” to describe the resultant Tax Code as a function of language, structure, and interconnectedness.

The objective of this chapter is to review prior work related to the composition of the complex technical structure. Emphasis is placed on the roles of cross-references in the formation of the Tax Code, and the reading and comprehension challenges faced by users who need to consult the Tax Code. It will be seen that there is a continuing and expanding interest in cross-references as an object of study, both within the context of legislative writing in general, and within the context of the Tax Code in particular. Throughout the chapter, claims are presented that the use of cross-references to enact elements of the legislative writing style and elements of the complex technical structure, represents a principal source of the complexity as the term was used in Chapter 1; i.e., challenging and difficult to read, to understand, and to apply. The chapter is organized around four main themes:



The Nature of Legislation Expression for Income Tax Law identifies four principal attributes of the legislative writing style and discusses the hybrid hierarchical-network information structure that serves as the foundation for the complex technical structure.

Linguistic Studies includes the topics of linguistic analysis of the legislative language and linguistic analysis of enacted legislation. The former topic is focused on the identification of multi-word patterns in legal writing synonymously identified as bundles, or n-grams. These word patterns, or cue phrases, are a core element of legal grammar in general, and of cross-reference grammar in particular. The latter topic includes what remains as the most detailed linguistic and structural analysis of enacted legislation, as well as an expanding body of computational work aimed at automatically detecting cross-references and developing semantic cross-reference classifications.

Network Centric Analysis of Legislation describes two recent projects aimed at determining, on a relative basis, the complexity of the laws contained in the United States Code (U.S.C). Results are included for the Tax Code. In both cases, the data can be interpreted as ranking the Tax Code as the most complex of all federal laws included in the U.S.C.

Related Human Subject Experimentation is a review of diverse experiments where each addresses an area of interest within the context of the present research. Included is a recent and important study in the context of the psychology of text comprehension. The results provide support for a claim that the reading of cross-references embedded in the text of income tax law will generate reading interruptions,

and that these interruptions in turn are a principal factor for observed reading and comprehension difficulties.

The following four sections explore and examine these four organizing themes. Thereafter, a Chapter summary and conclusion provides context for the research approach developed in Chapter 3.

### **The Nature of Legislative Expression for Income Tax Law**

This section identifies and discusses four principal attributes of the legislative writing style as used in drafting income tax law. A source of reading and comprehension difficulty is ascribed to each attribute.

#### *Pervasive use of cross-references*

The pervasive use of cross-references in drafting legislation is perhaps the most unique characteristic of the legislative writing style. Draftspersons employ cross-references for two purposes. One purpose is to achieve writing efficiency and consistency by referencing provisions in lieu of replicating the legislative text. Referencing definitions is the most common example of avoiding replication and ensuring consistency. A second purpose is for the draftspersons to make explicit and convey to readers their deep knowledge about the dependencies and interrelationships that exist among provisions within a given statute, and in some cases across related statutes. Collectively, the cross-references create, especially in the case of Tax Code, a vast network of interconnected statutory provisions. This network is a principal component of the Complex Technical Structure. Hence, as will be discussed throughout this chapter, the presence of cross-references in the Tax Code may negatively influence comprehension of the Tax Code in several ways.

*Elaboration*

The legislative draftsman is responsible for transforming the intent of the legislature into written statutory law. In the case of income tax legislation, the draftsman is faced with the daunting task of crafting rules sufficient in numbers and scope to capture the totality of potential financial transactions conducted by individuals and by economic entities. In doing so, attempts are made “to consider every possible eventuality, every possible instance of the proscribed behaviour that can conceivably occur” (Maley, 1987, p. 35). An income tax statute then will attempt to address this totality of transactions by the inclusion of numerous, detailed rules to capture the discerning shades of facts and circumstances. Miller (1993) characterizes this attempt to achieve complete legislative expression as the elaboration approach to rule making, where the rules themselves are intended to be a closed system, sufficient to determine proper rule selection and application in all circumstances. The result is not only lengthy rules, but also rules that are intricately interdependent (p.17). As illustrated in Chapter 1, this interdependence generally is formalized by the draftsmen’s use of cross-references. The characteristics and numbers of related rules can impose a considerable information processing burden on a reader (p.12).

The elaboration approach is due in large measure because statutes themselves generally cannot incorporate explanatory, illustrative, or interpretive elaborations. While such elaborations could well serve as useful comprehension aids, their inclusion would violate the long-standing principle that statutes present rules and nothing else (Lötscher, 2008, pp.133-135). This information processing burden referred to by Miller (1993) is rendered more problematical as over time, the number of rules and their degree of

interconnection tend to increase, as income tax legislation is adjusted to respond to economic and social concerns, and to remove opportunities for abuse. Flesch (1979b) describes how, over a period of several years, draftspersons used cross-references in an attempt to adroitly address taxpayer abusive behavior related to reducing tax liability by shifting assets to family members as Congress from time to time changed the definition of family member.

While draftspersons pursue the goal of constructing a set of complete rules, they also seek to use language in the most precise manner possible, so that the statutory text is free of ambiguity. Precision and absence of ambiguity are critically important because “a statute is a highly serious social document of ultimate authority which needs careful expression” (Trosborg, 1997, p. 30). According to Bhatia (1983), the complex forms of syntactic constructions and referential structures found in legislative text reflect how the draftsman has uniquely used the legislative language to meet the expectations of precision and absence of ambiguity. Bhatia goes on to say that such careful and detailed expression is not without the cost “of producing a [statutory] document which may be regarded by the reader as ... and unnecessarily complex, and hence unreadable” (Bhatia, 1983, p. 24). It is an interesting result that the customary legislative writing style, designed to meet the lofty objectives of completeness, precision, and absence of ambiguity, is one principal source of complexity associated with the complex technical structure.

### *Defined terms*

Morrison (2014) notes that the inclusion of defined terms, as well as cross-references, are essentially inevitable in a rules-driven, complex regulatory or statutory

scheme such as income tax law. The need to include defined terms is a direct result of using specialized terminology to express the rules that comprise legislative provisions. For certain terms, the volume of detail a reader must absorb can be considerable, and recursive-like processing of related nested terms and nested cross-references is thought to add to the reader's burden.

Morrison describes the definition of a qualified retirement plan to illustrate the overwhelming amount of information that comprises the definition (pp. 7-9). (The presence of the term *qualified* typically means tax favored). Tax Code section 401, Qualified Pension, Profit-Sharing, and Stock Bonus Plans, presents the definition of *qualified* in subsection (401)(a), Requirements for Qualification. Recall the provision hierarchy is section, subsection, paragraph, subparagraph, clause, and subclause. Subsection (401)(a) is subdivided into 37 paragraphs. Morrison's review of just paragraphs 1 through 8 identified 12 cross-references, and several defined terms. The 12 cross-references referred to provisions in six different Tax Code sections. Depending on the goal and prior knowledge of a reader, some number of the six sections would be accessed and read, and the reader would likely encounter yet more cross-references. For unfamiliar terms, a reader would follow a cross-reference to the provision containing the definition, or if no cross-reference were provided, would search for the definition. A full review of section 401(a) finds 93 cross-references to 26 different Tax Code sections.

### *Qualification*

Provisions are the principal legislative organizing unit. Two types of rules are contained in provisions, primary and secondary. Primary rules are expressed in relatively general terms, and secondary rules add details to address specific applications of the

primary rules. Hence, the expression of a specific point of law is a union of primary and secondary rules (Hart, 1994). Bhatia (1983) uses the parallel terms “main provisions” and “qualifying provisions” to describe the respective roles of provisions. The following Tax Code excerpt illustrates the relationship between a main and qualifying provision [annotation added].

26 U.S.C. § 1 – Tax imposed

(g) Certain unearned income of children taxed as if parent’s income

(1) In general [main provision]

In the case of any child to whom this subsection applies, ...

(2) Child to whom subsection applies [qualifying provisions]

This subsection shall apply to any child for any taxable year if—

(A) such child—

Paragraph (1), the main provision, states the general rule regarding taxation of certain unearned income of children. Paragraph (2), the qualifying provision, states the conditions under which the general rule will apply. In addition to specifying conditions, other types of qualification types include definitions, exceptions, limitations, ordering, and exclusions. Frequently, a qualifying provision will reference another Tax Code provision located in a different Code section via a cross-reference(s). Subclause (II), found in the full text of subparagraph (A) above, illustrates several types of qualification rules, here shown underlined.

(II) whose earned income (as defined in section 911(d)(2)) for such taxable year does not exceed one-half of the amount of the individual’s support (within the meaning of section 152(c)(1)(D) after the application of section 152(f)(5) (without regard to subparagraph (A) thereof)) for such taxable year. [§1(g)(2)(A)(ii)(II)]

In subclause (II), example qualification rules illustrate the qualification types, in reading order, of definition (two instances: as defined in; within the meaning of), ordering, and exclusion, respectively. Qualifying provisions require a reader to acquire

and to apply additional information in varying amounts and in varying ways depending on the construction of the qualification, and hence may contribute to readability difficulties.

### **Information Structures for Legislative Provisions**

Two different information structures are used to organize and to express the provisions that collectively comprise a statute. One information structure is hierarchy, and the other information structure is network. In this section, the contrasting features of the two information structures are presented, together with their hypothesized differential impacts on a user of the Tax Code. Due to the complexity of legislative rules, and the constraint that precludes augmenting the expression of rules with explanatory, illustrative, or interpretive information, Löttscher (2008) sets forth an imperative to render the logical structure of provisions as visible and transparent as possible. So doing will help to reduce impediments to comprehension of legislative writings. For the Tax Code (Title 26) and all other Titles included in the United States Code, the process of codification is motivated by achieving consistent and predictable expression of the provisions of federal statutes. Example approaches for achieving visibility and transparency are use of hierarchy together with consistently applied typographical conventions, and presentation of different categories of provisions in a reasonably consistent order.

#### *Hierarchy as an information structure*

All of a statute's provisions are explicitly organized and presented in hierarchical format. It is relatively straightforward to achieve logical visibility and transparency for the contents of hierarchical structures. For example, the typographical conventions

described in Chapter I render the global hierarchy of Title 26 visible at the table of contents level, and also render the local hierarchies that organize the text of provisions located at the section level and below the section level visible as well. The entire structure is transparent; hence, a reader will see a well-structured, ordered, cohesive set of provisions at the time a section is accessed.

In the absence of cross-references embedded in the hierarchy, hierarchy provides a natural, sequential, or linear, reading environment. In such environments, a sentence is read sequentially, word by word. At sentence end, the next sentence in physical order is read, and so on. In addition, and importantly, the provisions and their constituent rules in a given hierarchical branch will (mostly) address one income tax topic, although in certain cases, the topic will be multidimensional.

#### *Network as an information structure*

Network structures are used to organize the provisions that are external to any one section contained in a hierarchical branch. The legislative draftsman explicitly expresses qualification relationships between sections of the legislative text by cross-references directly in and across the body of the hierarchies containing the statute's legislative provisions. Hiltunen (2012) characterizes such relationship structures as the "web-like [network] nature of legislative texts" (p. 49).

In contrast to the visible and transparent hierarchical structures, the qualification network is invisible and not transparent. There is no visibility into the network formed by all Tax Code cross-references. At the time a reader encounters a cross-reference, the reader will not be able to see a complete network structure analogous to seeing a



complete hierarchical structure. In fact, for most cases, the reader will not be able to even visualize a partial network structure.

Different from a hierarchical information structure, a network information structure provides an unnatural, interrupt-driven, non-linear reading environment. In many cases, the complete order of provision reading will not be knowable by the reader, as there is no visibility beyond the referenced section until that section is accessed and read. In addition, unlike the one-topic content of a hierarchical branch, a reader may encounter several different income tax topics that will require the reader to engage in integrative rule processing, a potentially complex and difficult task.

This state of affairs can be problematic for a reader, since the qualification network is a source of important, additional statutory information. As Cyril (2013) explains:

“... the existence of formal [cross] references in the text of law semantically organizes the entire text as it extracts information from the whole text, which the author intends to operate in closer semantic or functional relationships” ... and that “relationships between the provisions ... may result in additional information that cannot be inferred from each of the provisions themselves, and which often determines ...the coherence of the whole text or fragment thereof” (pp.183-184).

There are two contextual dimensions to Cyril’s observation that reach respectively to the work of the draftsman and the task of the reader. Over periods of time, draftsmen define the semantic organization of the legislative text by adding, revising, and (rarely) removing cross-references from the statutory text. Even so, at any given point in time, the collection of cross-references is static. However, for a reader directed to one or several regions of the qualification network, the semantic organization is enacted dynamically, and is not determinable in advance. The dynamic enactment will be a function of a

reader's goal, and the reader's entry point to the statutory text. Consequently, and to varying degrees, perhaps depending on experience, a reader, a priori, will be unaware of additional information embedded in the qualification network until discovered and extracted when reading a networked structure of provisions. Discovery of unexpected information may disrupt the readers previously developed and stabilized understanding of the situation at hand.

#### *Information structure interactions*

The interactions of the hierarchical structure and the network structure create a distributed information structure for organizing legislative text. In this information structure, provisions are scattered throughout the hierarchy and related provisions are tangled together as a consequence of the networks created by aggregations of cross-references. Since "legal provisions take effect in the context of a number of other provisions and subprovisions" (Bhatia, 1983, p. 212), one consequence of the scattering and tangling is that "that text must be interpreted in the context of something expressed elsewhere" (emphasis added) (Bhatia, 1998, section Facilitating Textual Mapping). Hence, a reader's attention will be directed elsewhere to new information that presented in a context different from the reader's immediate context. Directing the reader's attention elsewhere may disrupt the reader's understanding developed up to the time of redirection (Miller & Kintsch, 1980). In turn, text comprehension may be negatively impacted.

#### *Summary*

This section identified four principal attributes of the legislative writing style as used in drafting income tax law: cross-reference, elaboration, defined terms, and

qualification. Each is a source of additional information that a reader will need to process, comprehend, and integrate, and then apply to a specific set of circumstances. All four sources contribute varying types of complexity to the complex technical structure.

A reader of income tax legislation will be confronted with the task of extracting, integrating, or synthesizing, information from two different and interacting information models, one that explicitly organizes provisions in well-formed, static hierarchies, and the other that implicitly organizes provisions in ill-formed, dynamically enacted networks. In such a hybrid information model, statutory provisions will be scattered throughout the hierarchy and simultaneously tangled together by cross-reference networks. Due to tangling, a reader's attention may be redirected from a provision in a specific local context to a related provision but in a different local context. In such a reading environment, a user may encounter difficulty maintaining the meaning of the statutory text, and as a result, comprehension may be impaired.

### **Linguistic Analysis of Legislative Language**

#### Lexical bundles

A lexical bundle, or n-gram, is a frequently occurring series of words. The concept of lexical bundle is not new, collocation being the original term. Hyland (2008) provides a history of interest in collocations and bundles. In his report of findings from his bundle analysis of academic writings in four disciplines, Hyland pointed to other researchers who have argued that multi-word patterns are core elements of grammars (p. 6). He continued, noting bundles are increasingly being seen "as important building blocks of coherent discourse and characteristic features of language use in particular

settings” (p. 8). Examples of bundles already presented in this chapter include as defined in, within the meaning of, after the application of, and without regard to.

A major motivation for conducting bundle analysis is to discriminate text-based genres by identifying bundles, the relative frequency of their occurrence, and their use patterns. There are two reports of bundle analysis that included legislative text. Both studies focus on 4-grams, or 4-word bundles. Gozdz-Roszkowski (2011) conducted a bundle analysis of the American Law Corpus (ALC), a corpus he developed of approximately 5.6 million words that includes 60 legislative U.S. documents comprising a total of 1.18 million words. (p. 27). Seven legal genres are represented in the ALC: legislation, contracts, judicial opinions, briefs, law textbooks, academic journals, and professional articles. For the legislative genre, 216 different bundles were identified. The sum of the number of instances of each 4-word bundle was 27,879 (a total of 115,516 words) which then represented 9.4% of the 1.18 million words in the legislative collection (p. 111). Such a high percentage supports the arguments reported in Hyland (2008) above that bundles are core grammatical elements in languages used for specific purposes.

Similar to Gozdz-Roszkowski (2011), Breeze (2013) studied the relative frequencies of 4-word bundles in four genres of legal English: legislation, academic law articles, case law (court opinions and law reports), and legal documents. At the conclusion to her study, Breeze notes that collectively, many bundles in legislation (and in legal documents) serve to “articulate the intertextual network of concepts” by explicitly orienting the reader [via cross-references] to other parts of the statute or to different statutes (p. 251).

### *Frame and slot model*

Breeze's (2013) orientation may be considered an instance of the frame and slot model (Biber, 2006, p. 162). In the context of the study, the frame and the associated slot represent a two-argument, composite syntactical unit for expressing cross-references in the text. The frame contains a bundle(s) that is intended to give a reader a context for the information in the referenced text unit, and in certain cases, to provide a reader with processing instructions that reflect the formal relationship between individual provisions in the law as recognized by the legislative draftsman (e.g., after the application of section ...). The slot is the argument that contains the object of the frame. For a cross-reference, the slot will contain the hierarchical address of the referenced text unit (e.g., section 911(d)(2)). In the context of coherence literature, a cross-reference frame would be considered as an explicit cue phrase, a marker placed in the text to signal a relation (Knott & Dale, 1994; Knott & Mellish, 1996).

### **Linguistic analysis of legislation**

#### *Classical analysis*

There are several reports of linguistic-oriented analysis applied to a body of legislation. Gustafsson (1975) conducted a syntactic analysis of the [British] Courts Act 1971. The motivation for her work was "... an attempt to discover if there are some specific syntactic properties which contribute to the alleged complexity of the language of the law" (Abstract, p. 5). Gustafsson's study, similar to Crystal and Davy (1969), focuses on the sentence as the primary unit of analysis, although unlike Crystal and Davy's qualitative analysis, her work is based in large measure on statistical profiles of syntactic elements. The principal elements that Gustafsson studied were sentence length

in words, numbers and types of clauses and their positions in sentences (beginning, medial, or end), clausal embedding, nominalized verb forms, and “an abundance of prepositional phrases.” She noted that two of these elements, nominalized verb forms and prepositional phrases, can be used to create “enormously complicated [grammatical] structures” (p.26). The linguistic bundles previously discussed are, for the most part, prepositional phrases, which in turn are the natural language expressions found in the frame of the cross-reference frame and slot model. To some degree, Gustafsson’s enormously complicated structures likely contain multiple cross-references. For example:

Without prejudice to the provisions of section 4(10) of this Act, directions under subsection (2) of this section may be given on behalf of the Crown Court by an officer of the Crown Court, but the power to make orders conferred on the Crown Court by subsection (4)(b) above shall be exercisable only by a judge of the court. (Courts Act 1971, Part II, Section 7(5))

In his detailed linguistic analysis of a UK statute, Bhatia (1983) presented the results of his analysis related to the physical positioning of qualifications within provisional sentences. A principal finding of interest here is that in many cases, the normal sequence of syntactic units in a sentence is “rendered discontinuous by embedded qualificational insertions” (p. 239), thereby creating a “syntactic discontinuity” that arguably will negatively impact a reader’s comprehension of legislative writing. In the context of the Tax Code, the embedded qualificational insertions will sometimes be cross-references. Syntactic discontinuity is a principal finding of Bhatia’s work, and the hypothesized negative impact on comprehension was supported in a human subjects experiment conducted as part of the study (Chapter Six, *Legislative Writing and the Reader*, pp. 229-240).

Bhatia's dissertation (1983) remains the most detailed linguistic and structural analysis of enacted legislation, in this case the (British) Housing Act 1980 that is comprised of approximately 100,000 words organized into nine parts and 155 sections. The principal motivation for the work was to identify sources of complexity in legislative writing to determine how best to help post-graduate students, whose native language was other than English, understand legislation written in English. Hence, there is an underlying concern throughout the work about the comprehension of legislative writing. Bhatia's approach to this task was to accurately and fully recover the intentions and thoughts of the legislative writer at work. To support Bhatia's analysis, access to a specialist informant (subject matter expert) was a component of the research methodology. The specialist informant, who had been a principal author of the British Housing Act, was available prior to and throughout the study to answer questions, offer suggestions, and to validate findings on request. Operationally, the study was organized around the question of how and why legislative documents are written as they are. Bhatia's analysis focused on identifying and categorizing the patterns of legislative language whose functions were to add qualifications to main statutory provisions. Qualifications are ubiquitous in statutes, as they shape the main provisions via elaboration and qualification to address immediate and foreseeable circumstances. Bhatia identified 10 qualification types without a claim to exhaustiveness. The qualification types were aggregated into three groups: "preparatory qualifications ... that specify circumstances to which the provision is meant to apply; operational qualifications ... that specify the manner in which the provision is required to operate; and referential qualifications ... to establish and specify a legal relationship between a provision in

which they occur and some other which may have legal bearing on it” (p. 63). Such relationships may change due to revisions in existing provisions or due to the addition of a new provision. Bhatia quotes the specialist informant who notes the responsibility of the legislative writer to document these relationships:

“... very rarely is a new legislative provision entirely freestanding .... it is part of a jigsaw puzzle .... In passing a new provision you are merely bringing one more piece and so you have to acknowledge that what you are about to do may affect some other bit of the massive statute book ....” (1983, p. 172).

Referential qualification is the qualification grouping closest to the interests of the present study. As will be discussed later, the embedding of cross-references in the statutory text is the method used to express referential qualifications.

For each of the four referential qualification types, Bhatia presents a textual pattern schema, together with schema instance examples, that expresses the qualification (Chapter Five, Referential Qualifications, pp. 172-228). He considers most of the schema instances to be complex prepositional phrases. The resulting inventory of complex prepositional phrases and adverbial clauses, together with their frequencies of occurrence predates the more recent work reported above in application of lexical bundle analysis to legislative writing.

The following textual patterns, present in the U.S. Tax Code, are examples of Bhatia’s four referential qualification types. Brackets indicate text not always present in an instance.

1. Qualifications that provide legal authority.

{under; in accordance with} [the provisions of] {section; subsection} [of the]

2. Qualifications that provide terminological explanation.

{as defined in; within the meaning of} {section; subsection}



3. Qualifications that provide textual mapping (i.e., text-cohering).

{referred to in; specified in} {section, subsection} [above; below]

4. Qualifications that provide legal scope.

{subject to; notwithstanding} [the provisions of] {section; subsection}

More recently, the NLP and machine learning communities are applying computational methods to automate and scale the identification, extraction, resolution, and classification of cross-references embedded in legal text. This work has been motivated by concerns, voiced by the requirements engineering community, that software requirements engineers working in regulated domains would encounter difficulties in accurately and completely understanding the intents and impacts of cross-referenced provisions in statutes and associated regulations. Consequently, software systems designed to support organizational compliance policies and procedures might not provide fully capable and compliant operational support (Breux & Anton (2008); Hamdaqa & Hamou-Lhadj (2011); Maxwell, Anton, Swire, Riaz, & McCraw (2012)).

#### *NLP and machine learning studies*

For computational cross-reference analysis, the network structures formed by the embedded cross-references are considered to be legal citation graphs. In a legal citation graph, nodes represent statutes and statutory provisions, and edges represent the cross-referenced relationships that are present within and between the statutes and their provisions. The principal goals of the computational-based work are to automatically identify and extract a high percentage of the embedded cross-references, and then to classify the cross-referenced relationships.

Based on a manual examination of the of the 2013 Income Tax Law of Luxembourg (written in French), Adedjouma, Sabetzadeh and Briand (2014) developed a set of regular expressions for detecting instances of legal cross-references and defined a grammar for the cross-reference patterns. The regular expressions and grammar were applied to the Income Tax Law text using NLP tools, and the results were presented in a taxonomy of cross-reference expression types by frequency of occurrence in the text (p. 67). Thereafter, Sannier, Adedjouma, Sabetzadeh, and Briand (2015) extended the work of Adedjouma et al. (2014) to include an examination of four additional Luxembourg statutes (in French), as well as the French and English editions of the Personal Health Information Protection Act of Ontario, Canada. Finally, Sannier, Adedjouma, Sabetzadeh, and Briand (2016) then extended the scope of their prior work by applying machine learning to the cross-reference expressions to label the semantic intent of the cross-reference. The label set included {compliance, constraint, definition, delegation, exception, refinement, and {amendment by ...}}.

In work similar to Sannier et al. (2016), Sadeghian et al. (2018) used the term *predicate* to identify the lexical bundle, or n-gram that preceded the alphanumeric location identifier(s) (the case of following the location identifier is not noted in their work). Sedeghian et al. (2018) defined a predicate feature set comprised of syntactic and linguistic variables that were used to represent each predicate as a vector comprised of a set of feature values for each predicate token. Predicates were extracted from a dataset of 1,000 randomly selected provisions contained in the United States Code. The vectors served as inputs to the machine learning algorithms that assigned labels for categorizing the purpose served by each cross-reference. The label set, proposed as a gold standard

label set, contained {legal basis, authority, definition, example or illustration, exception, criterion, limitation, procedure, amended by/amended to} (section 5.2, Design the Golden Labels).

The cross-reference taxonomies appearing in the independent but concurrent work of Sannier et al. (2016) and Sadeghian et al. (2018) were based on merges of and extensions to the earlier taxonomies developed in Breaux and Anton (2008), Hamdaqa and Hamou-Lhadj (2011), and Maxwell, et al. (2012). These two taxonomies then reflect the most robust analysis to date regarding the semantic intent of legal cross-references. In addition, the variety of presentation formats exhibited by cross-reference expressions has also been captured in prior work and extended by present work. For example, the robust grammar describing cross reference expression patterns in Adedjouma et al. (2014) was based in considerable part on the work of de Maat, Winkels, and van Engers (2006). The grammar captures several cross-reference expression types and an associated variety of cross-reference expression formats. The set of types includes internal/external, explicit/implicit, and simple/complex, where complex contains multivalued and multilayered. A multivalued cross-reference cites more than one provision by using AND/OR provision enumeration or provision ranges, or both. A multilayered cross-reference describes the full hierarchical path to a subprovision. (Adedjouma et al. (2014), pp. 66-67). A multilayered cross-reference can have multivalued arguments.

Bowers (1989), writing in the context of legislative expression, noted that accessibility to legislative language (or comprehension of legislative language) is a function of the reader's predictability of what types of information will be read [next] expressed in regular and uniform structures (p. 343). Applying Bower's perspective, and

given the variety of legal cross-reference semantics, and the variety of lexical bundles and cross-reference formats identified here, cross-references will be a source of readability complexity.

*Comparison with Bhatia (1983)*

Circa 2006, linguistic studies, focused on the cross-reference semantics of legislative and regulatory writings, began to appear in the literature with increasing frequency. Continuing to the present time, each new publication tended to review and compare prior semantic taxonomies to highlight an improved taxonomy being presented by the new work. Given the motivation for these recent works, it is not surprising that none include reference to Bhatia's work (1983). Still, in terms of the enduring nature of legislative (and regulatory) expression, some degree of overlap between Bhatia's cross-reference semantic taxonomy and the post-2005 taxonomies is a reasonable expectation. Indeed, there is substantial overlap between the post-2005 exception, constraint, and limitation cross-reference categories, and Bhatia's legal scope cross-reference. Legal scope considers the impact of one provision on the operation of another provision where the two provisions are related by a cross-reference(s). The Tax Code includes numerous provisions labeled Exceptions, Limitations, and Special Rules, all of which address legal scope. An examination of the provisions so labeled reveals that in the aggregate, they include expressions of rules that are encumbered with those elements of the statutory language and of drafting practices that many have considered problematical in terms of readability. The encumbering elements include conditional expressions, negations, chains of cross-references, nested provisions, and deep detail.

This concentration of presumed problematic elements in one of Bhatia's four categories, and their frequent presence in certain categories of Tax Code provisions noted above, suggests that for the study, two categorizations of cross-references will be useful in isolating the impacts of cross-references on readability. One category is cross-references whose purpose is providing legal scope. The other category, arbitrarily named support cross-references, is comprised of Bhatia's other three cross-reference types: providing legal authority, providing terminology explanation, and providing textual mapping.

### *Summary*

Over time, cross-reference pattern formalisms have been developed by the linguistic and NLP communities, as the importance of cross-references as a core element in legislative writings has become more evident. An early and independent work used simple schemas to define cross-reference patterns. Later, linguistics developed a frame and slot model to formalize the expression of n-grams, or bundles as the frame values, and the slot as the referred-to object of the frame. Frame values are typically complex prepositional phrases and are used to convey context to the reader. More recently, NLP researchers have used regular expressions and created formal grammars to describe cross-references, with the goal of developing cross-reference semantic taxonomies.

### **Network Analysis of Legislation**

The legal citation graph perspective, where nodes represent statutory provisions and edges represent the cross-referenced relationships between provisions, has motivated several investigations in addition to those previously described in the section NLP and Machine Learning Studies above. Two studies have been reported that include

complexity measures of the Titles that comprise the U.S.C. In both cases, the results can be interpreted as assessing the Tax Code (Title 26) to be the most complex U.S.C title. Katz and Bommarito (2014) presented and initially applied a conceptual framework and a multi-component measure designed to determine the relative rank ordering of the reading and comprehension complexity of each U.S.C. Title. While the Title is their unit of comparison, detailed analysis was conducted at the section level and then the section results were aggregated to the corresponding Title levels. Their analytical perspective is based in part on the field of knowledge acquisition which “studies the protocols individuals use to acquire, store, and analyze information” (p.340). The authors present a knowledge acquisition protocol in the context of a user, or reader, about to access the U.S.C. as a three-stage process:

1. Select an initial element of the Code corresponding to a concept of interest [or, begin at an already determined starting point],
2. Recursively assimilate the content of all sub-elements, and
3. When a citation [cross-reference] is encountered, apply the protocol recursively to the cited element (p. 346).

Katz and Bommarito (2014) define assimilate to mean “the process of reading and understanding the actual text [of a Title]” (p. 352), and complexity as “the [cognitive] cost of carrying out the acquisition protocol” (p. 340). Hence, a rank ordering of relative Title complexity will be the rank ordering of assimilation cost, or comprehension difficulty. The multi-component complexity measure is derived from their three-element framework that includes the factors of structure, language, and interdependence. The final

measure, normalized for size, included two structure factors, and one factor each for language and interdependence:

#### Structure

*Average tokens per section.* A count of the number of strings (words and alphanumeric identifiers) in each section of a Title, averaged over the number of sections in the Title.

*Average node depth.* A measure of the average hierarchic depth of each node in a section of a Title, averaged over the number of sections in the Title.

#### Language

*Entropy.* A measure of the concept or topic variance in a Title.

#### Interdependence

*Net Flow.* A measure of between Title interdependence calculated as the difference between the number of cross-references directed to a Title from all other Titles and the number of cross-references emanating from Title to all other Titles. Each Title received a rank order score for each of the four components. The Title's complexity rank was calculated as the sum of the individual component ranks divided by four (equal weighting). The Katz and Bommarito complexity measure ranked Title 26 the 2<sup>nd</sup> most complex U.S.C. Title. However, and more interesting, a different result is arguably obtained upon closer examination of the composite measures shown in Table 3 (columns rearranged to match the order of factors above).

Table 3

## Comparative Complexity Ranking of Title 42 and Title 26

USC Title	Tokens/ Section	Node depth	Entropy	Net Flow	Composite score	Composite rank
42. Public Health & Welfare	8	10	2	2	5.5	1
26. Internal Revenue Code	2	1	29	7	9.75	2

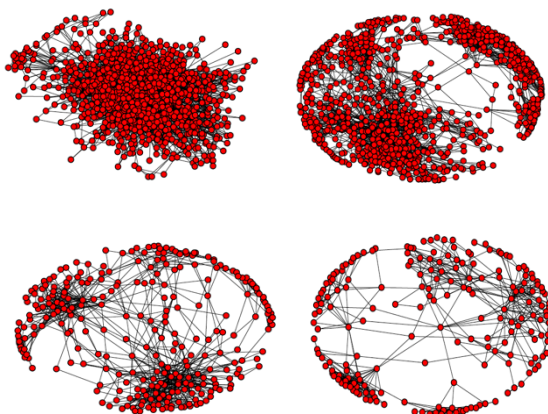
Derived from Katz and Bommarito (2014), Table 12, p.368.

In the Katz and Bommarito model, entropy is intended to measure a Title's concept variance by calculating Shannon entropy based on the frequency distribution of tokens within each Title. The higher the concept variance, the higher the assimilation cost, and hence a higher level of comprehension difficulty. Title 42 has essentially zero internal cohesion among its approximately 190 disparate chapters, or as the authors remark, "at closer inspection, the full Title [42] appears bound together by little more than the binding" (p. 358). The subtitle hierarchical level is not present in Title 42, arguably because of the lack of chapter relatedness. In comparison, Title 26 is comprised of approximately 70 chapters contained within 11 subtitles. Removal of the entropy factor then yields a Title 26 composite score of 3.33  $((2+1+7)/3)$ , a Title 42 composite score of 6.67  $((8+10+2)/3)$ , and finally a Title 26 composite rank of 1.

Li, Azar, Larochele, Hill, & Lo (2015) conducted another network analysis of the U.S.C. designed in large part to identify existing complexity measures, and how these measures can be applied to assess the relative complexity of the U.S.C. Titles. One such measure is the largest strongly connected subgraph. In such a structure, there is a path between all pairs of nodes (sections) and there are no unconnected sections. This measure provides insight into the degree of section connectedness of each Title, where greater degrees of connectedness are equivalent to greater degrees of complexity. The authors



term the largest strongly connected subgraph as the Title’s core. Figure 3 shows network visualizations of the largest cores (Title 20 not shown). Table 4 shows the core density (core sections / total sections) corresponding to the visualizations. The relatively extreme density of the Title 26 core, ~50%, is consistent with the finding that 97% of the Title 26 cross-references are to sections within Title 26 (Katz and Bommarito, 2014, p.361), rendering the Tax Code highly self-referential and complex. Li et al. (2015) note the average size of the largest connected subgraph across all U.S.C. Title is 90 (rounded) and that very large connected subgraphs in the U.S.C. are rare (p. 346).



*Figure 3.* Visualization of U.S.C. Titles with largest cores. Top row (Title 26 - Internal Revenue Code; Title 42 – Public Health & Welfare). Bottom row (Title 12 – Banks and Banking; Title 49 – Transportation). Title 20 – Education not shown. Li (2016), Appendix D.

Table 4

Top Five U.S.C. Titles with Largest Cores

Title	Core sections	Total sections	Core density
26	1,037	2,088	49.7 %
42	873	7,400	11.8 %
12	279	1,808	15.4 %
20	234	2,232	10.5 %
49	200	1,185	16.9 %

Section counts developed by Katz and Bommarito (2014).

Available at <https://github.com/mjbommar/us-code-complexity>

## **Related Human Subject Experimentation**

There are several reports of human subject experimentation that incorporated cross-references in the study design or reported results that included a cross-reference related finding. The reported results of each of the three quite different studies described here support, to varying degrees, a hypothesis that cross-references are a significant source of Tax Code complexity.

### *Tax problem solving*

Karlinsky and Koch (1987) conducted an experiment to determine whether readability complexity impacts the tax problem solving performance of tax practitioners. In this study, readability complexity is a function of presentation style. Two presentation styles were used in the study. One was the Tax Code statutory language, and the other was a commercial reporting service commentary. The latter included the statutory language accompanied by annotations, interpretations, and guidance. The task was to answer 15 textbook type questions about one of two Tax Code sections. The Tax Code sections were selected based on the author's assessment of the section's relative complexity. Section 179, Election to Expense Depreciable Business Assets, was determined to be less complex than Section 318, Constructive Ownership of Stock. The basis of their assessment was "Section 179 was chosen because it is relatively simple and self-contained (i.e., there is little cross-referencing to other Code sections. ... Section 318 was chosen because it is more complex in both length and content than Section 179. Section 318 ... may be viewed as a surrogate for eight major constructive rules, as well as a myriad of variations" (p. 26). Indeed, section 318(b) includes cross-references to eight different Tax Code sections.

The results showed that on average, subject performance, as measured by number of correct responses, was significantly better on the Section 179 (12.4 and 11.5 correct responses out of 15 questions for the Tax Code and Commercial Commentary presentations respectively, compared to the Section 318 task performance (6.6 and 7.7 correct responses out of 15 questions for the Tax Code and Commercial Commentary presentations respectively). Viewed as an absolute result, the author's commented that the tax practitioner performance on the Section 318 task was "disturbing" (p.32). In addition, on an authored developed seven-point Likert scale instrument, participants rated task difficulty more difficult for Section 318 compared to Section 179, and rated their task confidence lower for the Section 318 task compared to the Section 179 task. These relative results were independent of the two presentation styles. For this study, the contribution of the Section 318 cross-references to lower task performance is not determinable.

#### *Cross-reference categorization*

A second study focused directly on participant understanding of the role played by cross-references in healthcare legislation and regulation. Maxwell, Anton, and Earp (2013) conducted a study to determine whether individuals in the healthcare domain who normally participate in requirements activities such as healthcare professionals, legal domain experts, software requirements engineers, software architects, and software development managers, would find cross-referenced provisions problematical during the process of incorporating legal requirements into a software development project. Here, problematical means whether these individuals could correctly classify such provisions using a cross-reference taxonomy the authors had previously developed. The taxonomy

was comprised of 10 classifications: constraint [e.g., must include], exception [e.g., do not apply], definition, unrelated [to software requirements], incorrect [reference to a non-existent provision], general [reference to a non-specific law(s), e.g., any statute, regulation ...], and prioritization [of authority between new and existing law]. One classification, incorrect, was omitted from the study due to the additional effort required to render such a determination.

The 10 statutory and regulatory provisions used in the study were selected from four federal regulations and one federal statute. All provisions were related to information privacy and security matters. An expert group determined the correct classification for each provision; in some cases, a provision could correctly be assigned two classifications. The expert group consisted of the author, two senior PhD students, and a law professor who participated in the drafting of three of the federal regulations represented in the study.

Thirty-three participants, employed at electronic health record system vendors and in the broader Healthcare IT field, completed the online classification task. The task was designed to be brief. On average, participants completed the task in about eight minutes, after spending about two minutes reviewing the provided tutorial which contained several classification examples. Results were mixed, and not statistically significant. Participants performed best on assigning provisions to the exception and definition classifications, then on provision assignment to the prioritization and again exception classifications, and finally worst on provisions whose correct primary classification were constraint, unrelated, general, and again prioritization. The authors concluded that the study results provided initial and preliminary confirmation that cross-referenced provisions will be

problematical, to varying degrees depending on experience (in this study, greater experience was a mitigating factor) for individuals engaged in detailed analysis of text in regulated domains. The researchers noted an intention to replicate their study after addressing a few methodological issues such as providing an improved tutorial experience with feedback, and providing a software specification for context to address potential ambiguity for the unrelated cross-reference category.

#### *Sources of legislative reading difficulty*

The strategic goal of a third research study was to increase citizen access to the law by improving the readability of statutes and regulations (Curtotti et al., 2015a). There were two tactical goals of the activity. The first tactical goal was to determine to what extent machine learning could predict whether readers of legal language would assess specific sentences as easy to read/understand or difficult to read/understand. The second tactical goal was to develop an initial set of writing guidelines, based on the study research results, that had the potential to increase the readability of statutes and regulations (Curtotti et al., 2015b). The study was a collaborative effort between the Australian National University and Cornell University's Legal Information Institute (LII). The LII maintains an online collection of U.S. federal law and federal regulations and similar state level items, as well as primary world legal materials. The principal study corpus was comprised of random sentences sampled from the United States Code (USC) and the U.S. Code of Federal Regulations (CFR).

Crowdsourcing was the research methodology employed for the data collection phase (Curtotti, et al., 2015a). For a period of three months, when a user accessed a certain page of the USC or the CFR, a pop-up question about one corpus sentence located

on that page was displayed. Two question formats were used. One format asked whether the user agreed with a statement that the sentence was easy to read, or was difficult to read. Possible responses were displayed as a Likert scale in multiple choice format. The other format asked the user to rate the readability of the sentence on a seven-point semantic differential scale, from very easy to very difficult. (The Cloze procedure, where the user is asked to fill-in a blank that appears after every N words, was a third question format). A variety of statistical methods were employed to analyze the collected data. A set of sentences labeled easy or labeled hard was the first phase output related to readability. Consistent with the study's strategic goal, a large volume of demographic data was also collected and analyzed.

Machine learning was used in the second phase of the study to attempt to identify significant language features associated with those sentences labeled as difficulty. In turn, such language features would then serve as a foundation for creating drafting heuristics aimed at improving the readability of laws and regulations. Two findings, perhaps related, are germane to this proposal. First, most sentences that included more than six prepositions were labeled as hard. Second, most sentences that included three or more cardinal numbers also were classified as hard. The authors interpreted the presence of cardinals as the presence of cross-references (Curtotti et al., 2015b, pp. 58-59).

#### *Non-linear and interrupted reading*

The two structures of the hybrid hierarchical/network information structure that are used to organize and present the provisions of legislative text require users to read in both a sequential processing mode and in a non-sequential, or non-linear, processing mode. A user generally will read collections of text sequentially, one complete sentence

after the next. In contrast, a user encountering cross-references embedded in legislation provisions will suffer reading interruptions resulting in a mode change from sequential reading to non-linear reading. As Hiltunen (2012) noted,

Of crucial importance is how sentences are constructed syntactically with regard to their information structure. Sentences where the linear flow of information is repeatedly interrupted are likely to be more difficult to process ... (p. 42).

Hiltunen's comment, offered in the context of embedded clauses, can be viewed as an instance of Bhatia's concept of syntactic discontinuity developed in the specific context of qualification networks. The comment, when considered in the broader context of reading interruptions that result from the interactions of the hierarchical and network structures used to express statutory provisions as previously discussed, raises the issue of the relationship(s) between reading interruptions and complexity. The following summarizes a recent and important paper centrally concerned with the psychology of interrupted reading.

Short-term memory (STM) and long-term memory (LTM) have been core elements of most information processing models of human memory (Atkinson & Shiffrin, 1968). STM is a temporary memory store with limited capacity, fast retrieval times, and fast information decay rates. LTM is a permanent memory store with very large capacity, and retrieval times and information decay rates slower than those in STM. Baddeley and Hitch (1974) set forth an expanded view of STM termed working memory (WM), or synonymously, short-term working memory (ST-WM)). WM subsumed STM with its traditional attributes, but as a set of three specialized STMs together with an executive control system.

Another advancement in human memory theory occurred when Ericsson and Kintsch (1995) proposed the concept of long-term working memory (LT-WM) that, for individuals skilled and practiced in a specific domain, served as an extension to ST-WM but with the permanent (or durable) storage characteristic of LTM. One motivation for presenting the LT-WM concept was to answer a question such as: “How can working memory based on temporary storage [ST-WM] account for the fact that skilled activities can be interrupted and later resumed without major effects on performance?” (Ericsson & Kintsch, 1995, p. 211). In their paper, Ericsson and Kintsch relied on prior experimental work by others that showed performance on recall and recognition tasks after reading text passages was not disrupted by the introduction of an unrelated, attention demanding task after participants had read a passage but before they were presented with recall or recognition tasks.

Foroughi, Werner, Barragan, and Bohm-Davis (2015) challenged Ericsson and Kintsch’s assertion that interruptions while reading will not diminish text comprehension. Foroughi et al. based their challenge on the argument that recall or recognition is a weak test of reading comprehension, and that a more robust determination of comprehension should be based on task performance that require connecting and synthesizing information from a text. They designed an experiment to test their challenge. Twenty-four college students read and answered questions on each of four passages used to assess reading comprehension on the SAT. Each passage consisted of four paragraphs. The interruption task was a series of math problems. Each participant read the four passages under two different experimental conditions. Under condition 1, the interruption task was displayed when the participant used the space bar to retrieve the next paragraph. Under



condition 2, there was no interruption task between paragraph retrievals. After a complete passage was read, participants answered eight questions about the passage.

The statistically significant results showed that all participants performed better in the no interruption condition compared to the interruption condition based on the average number of correct responses. In addition, no individual participant's performance was greater under the interruption condition compared to the no-interruption condition (p. 706). These results raised questions about the viability of the LT-WM concept, one that had been readily accepted since the 1995 paper by Ericsson and Kintsch. Delaney and Ericsson, (2016) published a challenge to the conclusions of Foroughi, Werner, Barragan, and Bohm-Davis (2015) in the form of a LT-WM based explanation of their findings. Nonetheless, the challengers did note Foroughi et al. (2015) is an important contribution, as the experimental design introduced a new, higher-order test of text comprehension, and the experimental results may have identified limitations of LT-WM theory.

In the context of the study and given that when a reader encounters a cross-reference a reading interruption will occur, the results of Foroughi et al. (2015) support a claim that the presence of cross-references in income tax law will create reading and comprehension difficulties.

### **Summary and conclusions**

The semantic and syntactic properties of legislative language, the embedded and invisible cross-reference network, and the scattering and tangling of legislative provisions within a hybrid hierarchical/network information structure combine to create a complex and difficult technical reading environment. Separately, the semantic and syntactic components of this reading environment have been examined in prior studies. Cross-

references, within the concept and context of bundles are considered a core element in the legislative language grammar (Hyland, 2008) but their use embodies linguistic features argued or found to have negative impacts on the comprehension of text. For example, cross-references are composed of complex prepositional or adverbial phrases that may be logically compounded, and serve to express complex conditionals, and exceptions to legislative rules via the use of negation. The variety of how the legislative grammar is used to express cross-references can impose troublesome conditions of information unpredictability and unexpectedness on a reader (Bowers, 1989). Cross-references embedded in complex sentence structures create problematic syntactic discontinuities that lead to readability problems (Bhatia, 1983). Finally, cross-references underlie the creation of hostile, legislative reading environments, since their presence transforms a pure linear and consecutive reading environment into a hybrid, non-linear, interrupt-driven, challenging reading environment.

Notwithstanding the negative readability impacts of cross-references, as a discourse feature in legislative writing, cross-references serve several, foundational roles. Cross-references help to semantically organize the legislative text (Bhatia, 1983; Cyrul, 2013; Breeze, 2013), create discourse coherence (Hyland, 2008), and provide paths to additional, relevant information (Biber, 2006; Gozdz-Roszkowski (2011)). More broadly, cross-references are a foundational element of the qualification network, and a principal, defining element of Surrey's (1969) complex technical structure.

The next chapter presents the development of a human subjects experiment designed to isolate and assess the effects of cross-references on Tax Code readability and Tax Code scenario problem solving. The effects of the two cross-reference types

identified in this chapter, Scope cross-references and Support cross-references are assessed separately. The motivation for the experiment is to acquire additional insights into the parameters that impact Tax Code readability and comprehension, and thereby contribute to the expanding literature that is focusing on the relationship between cross-references and comprehension of statutory and regulatory writings in general, and on the U.S. Tax Code in particular.

## **Chapter 3**

### **Methodology**

#### **Motivation**

Chapter 2 concluded with a summary narrative of scholarly work focused on the interaction of an individual with the text of statutory and regulatory writings in general, and with the text of the U.S. Tax Code in particular. The narrative reveals a relatively low historical level of experimental work compared to the levels of other forms of analysis applied to these same types of writings. Hence, there is an imbalance between the scope and scale of research findings that point to potential sources of statutory complexity, and available empirical data obtained from human subject studies to assess such findings. Additional empirical data will help to increase understanding about the parameters that impact problem solving performance in cases where the actual statutory language needs to be examined.

Prior work supports a hypothesis that cross-references embedded in the statutory language is a parameter that impacts a reader's ability to comprehend and apply provisions of the Tax Code. The objective of the study is to determine to what extent, if any, the presence of cross-references in the statutory text of the Tax Code negatively impacts the readability, comprehension, and application of Tax Code provisions.

#### **Overview of the Approach**

A human subjects experiment was conducted to provide insights into the hypothesized negative relationship between the presence of embedded cross-references in the statutory text of the Tax Code and the ability of individuals to understand the meaning of statutory provisions in the context of tax scenario problem solving. Student

participants pursuing a degree program in accounting that included a required introductory federal taxation course were presented with a tax scenario, associated questions, and the legislative text of the Tax Code sections that address scenario questions. No other reference materials were available to participants. Each question included a list of Tax Code sections and provisions that are on point (i.e., relevant) to the task of developing a question response. Answers were of the form of either Yes/No, or the indication of a specific dollar amount. In addition, to decrease the frequency of guessing the Yes/No type questions, participants were asked to provide brief written explanations, with reference to specific Tax Code provisions, of their answers to certain questions. These explanations were not formally analyzed for the purpose of the study experiment but are retained for future analysis.

### **Development of the Experimental Scenario**

Design and development of the experimental scenario was a challenging endeavor, marked by several false starts, the associated creation of several scenarios, and a change in dissertation goal from developing a complexity measure for Tax Code sections and provisions, to an emphasis on better understanding the types of cross-references and their respective purposes, and the impact of cross-references on problem solving performance in the income tax domain. In addition, due to Tax Code provision interrelatedness, any given change to the scenario in-process often required the need to make adjustments elsewhere.

Satisfaction of the following interrelated scenario design goals was a major source of development complexity.

1. Achieve a measure of structural variety in Tax Code sections so that the complex technical structure of the Tax Code was adequately represented. This necessitated the presence of at least one major (within the context of the scenario) Tax Code section where the prevailing information structure was predominately hierarchical, and at least one major Tax Code section where the prevailing information structure was predominately network.

2. Achieve a measure of cross-reference volume and variety so that both supporting cross-references and legal scope cross-references are meaningfully represented. This goal impacted the process of selecting candidate Tax Code sections for topic inclusion in the scenario and associated questions.

3. Design scenario questions to direct participant reading of Tax Code provisions to exploit the presence of cross-references to the greatest extent possible in the context of the study's goal and experimental hypotheses.

4. Maintain participant intellectual accessibility to the Tax Code provisions that participants will be directed to read. Since the research interest in the study is focused on the impact of cross-references on participant problem solving outcomes, and not on determining the extent of participant Tax Code knowledge, a decision was reached to limit scenario topics to those topics covered or to topics related to those topics covered in the typical undergraduate introduction to federal income taxation course.

5. Maintain sustained engagement with the experimental task. Topic selection and control of difficulty level were thought to be two methods of supporting sustained engagement.

Eventually, the idea of family emerged as an attractive tax domain on which to base the scenario. The material in introductory level federal income tax courses cover basic family taxation issues, and the Tax Code sections that are associated with basic family matters are largely free of esoteric, specialized tax terminology. These two factors, prior exposure and family domain, are considered to be supportive of scenario design goal four.

Satisfying scenario design goals one was accomplished with the identification of two core scenario sections, §2 and §152. These sections include the rules for making determinations about who may qualify for tax-favored head of household status, and who may qualify as a dependent, respectively. Section 2 is predominately a network structure, characterized by a relatively small number of nodes and where the preponderance of cross-references are external cross-references. In contrast, §152 is an example of a predominately hierarchical structure, one that is characterized by a relatively large number of nodes and one where the preponderance of cross-references are internal cross-references. The two sections include relatively complex logic as evidenced by the four textbook pages of decision trees in flowchart format that were required to document the rule flows and branching points for determining head of household status and qualifying as a dependent (Spilker et al., 2019, pp. 4.27 – 4.30).

The rules in §7703 differentiate between not being married in the legal sense, and not being married for purposes of the Tax Code. Not being married is one of several criteria that must be satisfied to claim head of household filing status. Scholarship taxation, covered in §117, was added as a likely area of student interest, and thus to help support scenario goal five. Section 21, the section that contains the rules for the tax

treatment of expenses for household and dependent care expenses necessary for gainful employment, and Section 213, the section that contains the rules for deductibility of medical expenses, were added because they contain several instances of legal scope cross-references.

Section 152(e)(1), is a provision that specifies the circumstances under which one parent's right to claim a dependent exemption can effectively be relinquished to the other parent. Such an action is common among parents who no longer live together. This provision was integrated into the scenario to add a measure of incremental and wide-impact complexity to the scenario.

In the aggregate, the seven Tax Code sections include 100 cross-reference instances (not all unique), 48 of type supporting and 52 of type legal scope. These cross-references support scenario design goals two and three.

Participant Scenario Materials are available in Appendix A. The set of supporting scenario Tax Code Sections is available in Appendix B.

### **Experimental design**

A within-subjects design using one scenario (two-parts) was selected as the experimental design type. This design will provide a larger sample size for each question than would otherwise be available with alternative designs that assign differing treatments to subsets of participants. For the study, use of a within-subjects design importantly avoids the substantial challenge of developing two or more scenarios of formally determined equivalent complexity that would be required with use of a between-subjects design.



### *Experimental controls*

The experiment was designed to determine task performance under conditions of a reader encountering varying numbers and types of cross-references embedded in the text of several Tax Code provisions. There are a variety of potential participant behaviors and task design factors that plausibly represent threats to internal validity that require controls for elimination or mitigation.

Participant search behavior. Individuals differ in Tax Code search efficiency and effectiveness (Barrick, 2001). A control is needed to eliminate participant search for sections and for provisions that are applicable to scenario questions. Here, an elimination control is to provide participants with only the Tax Code sections needed to respond to scenario questions.

Content and reading order. What provisions a participant will choose to read and in what order may affect task performance. A mitigating control is needed to constrain, at least initially, what provisions of the provided Tax Code sections participants will read, or reread, and in what order. Absolutely controlling what provisions will be read and in what order cannot realistically be accomplished. Participants may self-distract by browsing or reading non-essential provisions, or perhaps more likely, reread certain provisions. Here the mitigating control has two components: (1) the presentation sequence of questions, and (2) the inclusion, for each question, of guidance specifying what provision(s) need to be read. If a provision contains a cross-reference to another provision, there is a strong expectation that the participant will follow the cross-reference to the referred-to provision.

Prior knowledge. Participants may be able to develop correct answers to questions without reference to the provided Tax Code sections. All participants will have completed or will be completing a first undergraduate course in Federal income taxation that covers most of the topics presented in the scenario, although the course coverage did not necessarily cover the aspects of a given topic that are present in the scenario. In addition, the volume of detail that needs to be recalled to correctly respond to the scenario questions is considerable. Given the experimental condition that participants will have no access to any reference materials (except the relevant Tax Code sections) or to any tax preparation software, the likelihood of consistently developing correct questions responses without reference to the provided Tax Code sections is considered a low probability outcome. Finally, participants are asked to provide written explanations of their responses to several questions, with specific reference to Tax Code provisions. To the extent participants comply with the explanation requests, these requests will serve as a control that should largely mitigate participants developing responses but not relying on the provided Tax Code sections.

Information overload. A control is needed to mitigate the potential negative effects of information processing overload caused by the sheer number of sections and provisions a participant will be required to read. Here, the mitigating control is to design the scenario such that the number of sections and the number of provisions to be read are not overly burdensome from an information processing load perspective.

## **Experimental Variables**

*Dependent variable.*

Response to question<sub>i</sub> (correct=1, incorrect=0)

The determination of whether a participant can read, understand, and apply Tax Code provisions for problem solving is based on participant responses to questions so designed that the correct response is objectively determinable with access to relevant Tax Code sections, given prior or current exposure to one introductory course that covers personal Federal income tax.

*Independent variables.*

The selection of independent variables is based on discussions presented in the prior two chapters. In Chapter 1, a hybrid information structure was described in which is arranged the [large numbers of] Tax Code sections and their associated provisions. The hybrid structure is a hierarchical arrangement of sections and provisions on which is imposed a network structure comprised of cross-references. Each section, and each provision, is a hierarchical node. Independent variables IV1, IV2, and IV3, shown below, represent, for each scenario question, the volume of information that a participant will read, and the number of nodes and the number of cross-references the participant will experience while reading the applicable Tax Code sections. For the study, the totality of cross-references encountered is conceptualized as the number of reading interruptions a participant will encounter during the directed reading activity associated with a given question.

Independent variables IV4 and IV5, shown below, represent, for each scenario question, the classification of cross-references, discussed in Chapter 2, as supporting or legal scope cross-references. For experimental design purposes, this classification seeks to isolate the effects of the two cross-reference categories on the dependent variable, as it

is hypothesized that legal scope cross-references will have a greater negative impact on readability compared to the impact of supporting cross-references on readability.

The five independent variables defined for the study are as follows:

**IV1.** Total tokens read per question

**IV2.** Total number of nodes read per question

**IV3.** Total number of cross-references encountered per question

**IV4.** Total number of supporting cross-references encountered per question

**IV5.** Total number of legal scope cross-references encountered per question

The definition of each independent variable, together with how each independent variable is counted or categorized as a support cross-reference or a legal scope cross-reference, is discussed below.

**IV1.** Total tokens read per question.

A token is a contiguous string of alphanumeric characters terminated by a blank character. The motivation to include token counts is to capture, on a relative basis, the amount of information processing that will be encountered by a participant during the reading of Tax Code section(s) in the context of a given question. The token concept applies to individual words, individual amounts, hierarchical identifiers of sections and provisions, and cross-reference expressions. The Word Count feature in Microsoft Word is used to calculate token counts. Word Count is applied only to tokens associated with provisions participants are instructed to read.

**IV2.** Total number of nodes read per question

A node is a labeled section or provision where the label is a hierarchical identifier optionally followed by a text description. A node count for a provision, or for part(s) of a

provision, that a participant will read is determined according to the following rules:

- The node that represents the section name (top level of hierarchy) is not counted. This rule avoids the problem of multiple counting in the likely event that a question requires reading more than one provision of a given section.
- Node counting begins at the node that represents the beginning of a provision, or at the beginning of the hierarchical path leading to the provision, as specified by the reading instructions that accompanies each scenario question
- Node counting continues along the provision hierarchy until the logical (for a given question instruction) termination node is reached.
- Intermediate nodes that a participant will not be directed to read are excluded from the node count.

#### Node counting examples

Example 1.

#### **§152. Dependent defined**

##### **(a) In general**

For purposes of this subtitle, the term "dependent" means-

- (1) a qualifying child, or
- (2) a qualifying relative.

Participants are asked to read §152(a). Participants will read subsection 152(a), and paragraphs 152(a)(1) and 152(a)(2). Section 152 is the name of the section and is not counted. The node count is three.

Example 2.

#### **§117. Qualified scholarships**

##### **(d) Qualified tuition reduction**

###### **(1) In general**

Gross income shall not include any qualified tuition reduction.

**(2) Qualified tuition reduction**

For purposes of this subsection, the term "qualified tuition reduction" means the amount of any reduction in tuition provided to an employee of an organization described in section 170(b)(1)(A)(ii) for the education (below the graduate level) at such organization (or another organization described in section 170(b)(1)(A)(ii)) of-

(A) such employee, or

(B) any person treated as an employee (or whose use is treated as an employee use) under the rules of section 132(h).

**(3) Reduction must not discriminate in favor of highly compensated, etc.**

**(4) Repealed. Pub. L. 101-140, title II, §203(a)(1), (2), Nov. 8, 1989, 103 Stat. 830**

**(5) Special rules for teaching and research assistants**

In the case of the education of an individual who is a graduate student at an educational organization described in section 170(b)(1)(A)(ii) and who is engaged in teaching or research activities for such organization, paragraph (2) shall be applied as if it did not contain the phrase "(below the graduate level)".

Participants are asked to read §117 (d)(1), (d)(2), and d(5). Participants will read 117(d), 117(d)(1), 117(d)(2), 117(d)(2)(A), 117(d)(2)(B), and 117(d)(5). Section 117 is the name of the section and is not counted. There are eight nodes starting at 117(d) and ending at 117(d)(5). The node count is six, since the participant will be asked to read all nodes with the exception of the two intermediate nodes, 117(d)(3) [text removed for this example] and 117(d)(4). Node counting is a manual activity.

**IV3. Total number of cross-references encountered per question.**

Each occurrence of a cross-reference adds one to the total cross-reference count, and adds one to the supporting or legal scope cross-reference category, whichever is appropriate. In cases where a cross-reference expression includes more than one cross-reference (i.e., the cue phrase is followed or preceded by multiple cross-references), the cue phrase is distributed over all cross-references and so each cross-reference is considered a separate instance. For example, the cross-reference expression subsections (a) and (d) shall not apply is equivalent to the two cross-references subsection (a) shall not apply and subsection (d) shall not apply. In this case, the total cross-reference count

and the legal scope account are incremented by the value two. A similar treatment of determined without regard to subsections (b)(1), (b)(2), and (d)(1)(B) increments the same two counts by the value three.

Chains of cross-references are similarly counted. For example, Gold Scenario Q4 includes instructions for participants to read §117(d)(1) and (d)(2). §117(d)(2)(B) contains a cross-reference to §132(h) which in turn contains a cross-reference §152(f). Since there is a strong expectation that participants will follow the §132(h) and §152(f) cross-reference, these two cross-references occurrences add two to the total cross-reference count, and in this example, add two to the support cross-reference count.

There is one categorical exception to the counting rule, namely multiple occurrences of stipulated cross-references. Across the text of the six Tax Code sections that served as the participant reference material, there are multiple instances of cross-references to three provisions, each located in one of three additional Tax Code sections, that this researcher considers immaterial within the context of the study. In the instructions to participants, these cross-references will be described and identified as being stipulated, and hence the sections containing these sections are not provided in the section reference materials. For purposes of counting, the counts pertaining to the question where the first occurrence of each of the three stipulated cross-references was encountered will be incremented. The reasoning for adopting this counting rule is that on first occurrence, the participant will likely experience an interruption, if not for anything other than novelty.

The provisions stipulated in the instructions to participants were:

§170(b)(1)(A)(ii). This provision describes the term *educational organization*.

The description is straightforward and is essentially generic. Participants will experience a reference to §170(b)(1)(A)(ii) five times while reading §117, Qualified scholarships, and once while reading §152(f)(5), Special support test in case of students.

§151. This section provides the legal authority to deduct the exemption amount for each individual who qualifies as a dependent under §152. Participants will experience a reference to §151 twice while reading §2(b)(1), the general rule that defines the criteria for head of household filing status, and once, while reading §7703(b).

§6013. This section contains the rules for married couples to file joint returns. In the scenario, the filing of a joint return has negative consequences, but there is no need for participants to read the §6013 rules. Participants will experience a reference to §6013 twice while reading §152, Dependent defined.

#### **IV4. Total number of supporting cross-references encountered per question**

As in Bhatia (1983), the presence of specific cue phrases are the principal means used in the study to categorize cross-references. Examples of cue phrases associated with three supporting cross-reference subcategories identified in Chapter 1 [subcategories shown here in brackets] that are found in the study Tax Code sections include under the provisions of [Legal Authority]; as defined in and within the meaning of [Terminological Definition]; and for the purpose of [Textual Mapping].

Bhatia (1987) notes that the principal goal of textual mapping is to provide a text-cohering function to improve the readability of legislative texts (p. 2). The drafting strategy to address readability improvement is to avoid overloading the reader with detail at a given point in a provision by postponing some detail to a following provision.



Typically, the strategy is implemented with a pair of cross-references, one to postpone detail, and the other to remind the reader of the context of the postponed detail. Tax Code sections 152(d)(1)(A) and (d)(2) illustrate the strategy (underlining added).

**(d) Qualifying relative**

For purposes of this section-

**(1) In general**

The term "qualifying relative" means, with respect to any taxpayer for any taxable year, an individual-

(A) who bears a relationship to the taxpayer described in paragraph (2),

**(2) Relationship**

For purposes of paragraph (1)(A), an individual bears a relationship to the taxpayer described in this paragraph if ... is any one of the following ... :

(A) A child ...

(B) A brother, sister, ...

...

(H) An individual, ... who has the same principal place of abode of the taxpayer ...

This approach textually separates a relationship requirement, (d)(1)(A), from the specification of relationship types that will satisfy the requirement, (d)(2). Bhatia (1983) also illustrates the strategy using a below/above pair of cross-references (pp. 208-209).

**IV5.** Total number of legal scope cross-references encountered per question.

Two approaches were taken in the study to identify legal scope cross-references. One approach was reliance on certain frequently occurring Tax Code provision headings: **Exceptions, Limitations, and Special Rules**. Provisions and their embedded cross-references located under one these headings address legal scope. (In uncommon cases, no cross-reference will be present in such provisions).

The other approach is again, as noted in the IV4 discussion above, reliance on specific cue phrases but here the cue phrases of interest signal a legal scope relationship. Common examples of such cue phrases in the Tax Code are notwithstanding, determined without regard to, and in those cases where the drafting objective is to implement a legal

scope relationship rather than to provide a text-cohering relationship, for purposes of.

Other cue phrases present in the study Tax Code sections that signify legal scope include but for, but not if, except as provided in, and shall be applied as if. These cue phrases also are frequently present in provisions located under of the three headings shown above.

The general rule adopted in the study was to categorize every cross-reference that participates in a legal scope relationship as a legal scope cross-reference.

### **Independent variable value assignments**

The Tokens, Nodes, Total Cross-References (Tot\_CR), Support Cross-References (Support), and Scope Cross-References (Scope) independent variables are measured by counts. Tot\_CR (IV3), Support (IV4), and Scope (IV5) are normalized both by tokens and by nodes, and incorporated in the hypotheses as cross-reference densities.

Table C1 (Appendix C) shows the provisions that pertain to each question, together with counts for each of the independent variables. The densities are calculated from the Table C1 data by dividing the IV.3, IV.4, and IV.5 values by the token and node counts respectively. The density values are shown in Table C2.

### **Hypotheses**

There is one hypothesis for each of the five independent variables. In addition, there is an additional hypothesis that reflects the assertion that legal scope cross-references are especially problematical for users of the Tax Code.

**H1:** The number of **tokens** a participant is required to read while formulating a question response is predictive of whether the answer is correct or incorrect.

**H2:** The number of **nodes** a participant is required to read while formulating a question response is predictive of whether the answer is correct or incorrect.

**H3:** The density of **total cross-references** a participant will encounter while reading required provisions is predictive of whether the answer is correct or incorrect.

**H4:** The density of **Scope cross-references** a participant will encounter while reading required provisions is predictive of whether the answer is correct or incorrect.

**H5:** The density of **Support cross-references** a participant will encounter while reading required provisions is predictive of whether the answer is correct or incorrect.

**H6:** Questions that require participants to read provisions with a greater density of **Scope cross-references** will have a smaller number of correct responses compared to questions that require participants to read provisions with a greater density of **Support cross-references**.

#### **Data collection and data preparation**

Data will be collected from two participant source documents:

1. Scenario Question document (contains answers and explanations)
2. Participant study experience survey

Participant responses to scenario questions were manually entered into an Excel spreadsheet coded as 1=correct answer; 0=incorrect answer. A master dataset was then generated comprised of, for each participant, one data tuple per question; response, where each question data tuple contains the following elements:

<b>PID</b>	participant ID
<b>Site</b>	experimental site {1; 2 (2.uGrad), 3 (2.Grad)}
<b>Scenario</b>	{Blue, Gold}
<b>Q<sub>i</sub></b>	scenario question number;
<b>Q<sub>i</sub>.Response</b>	response for scenario question Q <sub>i</sub>
<b>Q<sub>i</sub>.Data tuple</b>	the data tuple associated with each Q <sub>i</sub> obtained from Table C2

The resulting dataset, after adjustments discussed in the following chapter, contained 969 data tuples that included 697 undergraduate student participant tuples, and 272 graduate student participant tuples. The participant explanations were scanned into PDF format and saved for analysis in a future study

The participant study experience survey is principally a modified version of the Reader Complexity Elicitation (RCE), a Likert-scale instrument developed in the context of human subjects research in the Tax Code domain (Koch and Karlinsky, 1984; Karlinsky and Koch, 1987). The general modification was to pluralize the term *passage* (a synonym for Tax Code section) to sections. (Subjects in the above referenced studies received only one Tax Code section). One item about task performance confidence was added after noting that Karlinsky and Koch (1987) apparently did the same, although they did not include the revised 1984 RCE in their paper.

The study experience survey data is not intended to be formally analyzed, but rather may be used informally to better understand the experimental task results, and also to suggest additional analysis of the results not previously considered.

### **Data analysis**

Regression is a robust statistical model well-suited for behavioral research (Cohen, Cohen, West, & Aiken, 2003). For the study, the ordinary regression model cannot be used since the dependent variable for each scenario question is dichotomous. Hence, the Binary Logistic Regression Model, a statistical model developed for use in cases where the dependent variable is dichotomous, was the statistical model used in the study. The Minitab Express for Mac statistical software system was used for regression modeling.

### *Data models*

As the independent variables were all hypothesized to contribute to problematic user readability experiences, the principal study interest was to determine to what extent did the independent variables actually contribute to readability problems.

The approach to data modeling and analysis used two sets of the same binary logistic models, one set of models for undergraduate participants and one set for graduate participants. As discussed in the following chapter, to minimize or eliminate collinearity effects, all the hypothesis testing models, except for Hypothesis 6, included only one independent variable. Since Hypothesis 6 is a comparison of two independent variables, the Hypothesis 6 regression models included two independent variables.

Hypothesis 1 is tested using the non-normalized Tokens independent variable, and Hypothesis 2 is tested using the non-normalized Nodes independent variable.

Hypotheses 3 through 5 are each tested with three models: a model using one independent variable and two models using .Tokens and .Nodes normalizations of that independent variable. The non-normalized variables are included in hypothesis testing to determine whether variable normalization results in more meaningful and better fit models. Hypothesis 3 is tested using Tot\_CR. Hypothesis 4 and Hypothesis 5 are designed for more granular analysis and insight, and so separately test the two components of Tot\_CR, Scope CR and Support CR.

Hypothesis 6 is designed to compare the relative effects of participants reading Scope CR and reading Support CR on their development of question answers.

### *Hypothesis testing*

The main objective in binary logistic regression is to determine how one or more independent variables affects the probability, or alternatively the odds, that an observation will be a member of one or the other of the dependent variable's two categories. For the study, an observation is represented by a question data tuple as noted above. The dependent variable's two categories are correct question response (=1) and incorrect question response (=0).

#### *Interpretation of results*

The sign of the beta coefficient in a binary logistic model indicates whether the outcome (in the study whether a question answer is correct) is more likely (positive coefficient) or less likely (negative coefficient). The exponential of each beta coefficient yields the odds ratio. The odds ratios informs about the relative magnitude of the effect associated with each independent, or predictor, variable.

### **Study Development Activities**

#### *Initial work*

Early in the Spring 2016 semester, the researcher reached out to a taxation course instructor at an institution that would become a study site to discuss the possibility of conducting a pilot study in an ongoing taxation course. After reviewing the researcher's draft pilot materials, the instructor provided suggestions for wording revisions to the scenario text and to the scenario questions. The pilot was conducted using the revised pilot materials in one 75-minute class session late in the same semester.

The results of the pilot were disappointing to the course instructor and to the researcher. Students apparently did not treat the exercise seriously as evidenced by mostly incorrect answers and minimal or no accompanying written explanations. In a

debrief meeting, the course instructor offered to make a similar class available in the future as a formal study site. In that meeting, a decision was reached to treat the experimental task as a graded, in-class exercise to increase the likelihood of more intentional student participation.

### *Refinement*

During academic years beginning Fall 2016, the researcher reworked the pilot scenario materials with the objective to provide a richer experimental task. The number of tax topics and subtopics were moderately increased which then enabled changes in the design of questions so that future study participants would encounter an increased number of cross-references while reading Tax Code sections. Concurrently, the researcher began development of a formal dissertation proposal.

In the Spring 2017 semester, the researcher met with the tax instructor to review the most recent version of the scenario materials. Upon review, the instructor recommended that the scenario be divided into two scenarios, and that each scenario be presented to a future class in two consecutive class sessions so that sufficient time would be available for participants to develop the required answer explanations. In addition, the instructor provided a small number of technical corrections and one question wording revision. Thereafter, the researcher began to develop documentation of the scenario solutions.

In early Summer 2018, an informal pilot study was conducted with three senior-level, high-GPA undergraduate students who had completed the introductory taxation course at the future site 1. The purposes of the pilot study were to (1) to debug the scenario and the associated questions, and to adjust these items as indicated, and (2) to

determine whether the experimental task is accessible to student participants and to adjust the scenario and question difficulty as indicated. There were three principal findings. First, while the students had encountered cross-references in an intermediate accounting course (no taxation topics), they were unfamiliar with the organization and naming of Tax Code sections. Second, in the scenario questions, there were some mismatches between the language used in a few questions and the language used in the Tax Code that caused a measure of confusion on the part of the students. Third, there were some provisions shown on the scenario question documents as to be read, but were not essential in the context of the study objective. These provisions included the mechanics of calculations, especially complex provisions, and provisions not needed to correctly respond to a question.

Based on these findings, the following refinements to the experimental materials were undertaken:

- A brief tutorial about the organization and naming of Tax Code sections was prepared (Appendix A) for study site course instructors for online distribution to their students one class session prior to the class session when the in-class exercise would begin.
- A few scenario questions were edited so that wherever possible, the questions incorporated matching Tax Code statutory language.
- For one scenario question, a stipulation was provided that informs a percentage value rather than requiring study participants to calculate the percentage.
- For one complex subprovision, a stipulation was placed in one question's reading directions to note that the subprovision's requirements are satisfied.



- For a few questions, the question instruction [but not] that signifies *do not read* was expanded to include non-essential provisions.

#### *Experimental sites*

Also in Summer 2018, the researcher followed up on a 2017 telephone conversation with the Chair of an accounting department to revisit the possibility for that institution to become a second study site. After an arranged meeting between a taxation course instructor and the researcher, an agreement was reached to conduct the study as a graded, in-class exercise at that institution.

#### *External review of scenario materials*

In Fall 2018, the scenario materials and researcher-developed scenario solutions were provided to a private practice CPA for a technical accuracy review. The review was completed in early November, 2018. One solution error was identified, and language edits were recommended to the documentation of several question answers. These changes were incorporated into the Scenarios Solutions document (Appendix E).

#### *Tax expert reviewers*

The study's tax expert reviewers are shown in Appendix F.

#### *Study Approvals*

IRB applications to the two study sites, and to NSU, were submitted and approved during October and November, 2018 (Appendix D). The NSU dissertation proposal was defended and approved in November 2018.

## Chapter 4

### Results

#### Conduct of the Experiment

The experiment was conducted at two Western New York college sites in mid-November 2018. Three groups of accounting major students comprised the study participant set: 25 undergraduate students at site 1, 38 undergraduate students at site 2, and 16 graduate students at site 2. All students were enrolled in tax accounting classes, and the respective classrooms served as the sites for administration of the experimental task. The undergraduate students were enrolled in introductory income tax courses where the principal focus is on taxation of individuals. The graduate students were enrolled in an advanced taxation course where the principal focus is on taxation of business entities.

#### *Schedule*

As recommended by both site course instructors, the experiment was conducted over two consecutive 75-minute class sessions. At site 1, the two sessions were a Tuesday and a Thursday. At site 2, the two sessions were a Wednesday and a Monday.

#### *Material distributed pre-exercise start*

Two days prior to the respective experiment start dates, the researcher provided the site 1 course instructor, and the site 2 course instructor (same instructor for both the undergraduate and graduate courses) a brief document (1.5 pages) entitled The Organization and Naming of Internal Revenue Code Sections. The instructors then posted this document to their respective course websites, and notified their students of the

availability of the material, mentioning that this material should be reviewed prior to the beginning of the in-class exercise.

*First class session*

At the beginning of the first in-class exercise day, the researcher was present in the classroom to obtain written informed consent. The course instructor introduced the researcher and then exited the classroom. The researcher distributed a consent document to each student, together with an attached index card that showed a unique, randomly generated alphabetic code. The researcher first noted that granting consent means the course instructor would anonymously share the student's exercise answer documents with the researcher. The researcher then reviewed each section of the consent document, explained how the random code would provide anonymity, and answered any questions. The protocol for student consent decision was then presented. If a student wished to consent, then that student would enter the code in the designated area of the first page of the consent form, sign the last page of the consent form, and retain the code for placement on the exercise documents. If a student decided not to consent, that student would enter their name on the first page of the consent form, not sign the document, and not retain the code.

Thereafter, the researcher stated he would exit the classroom, that the course instructor would shortly return, and that during this brief time interval students would make their participation decision, and place their consent document in an envelope at the front of the classroom. After all students submitted their consent documents, a student volunteer sealed the envelope and provided the envelope to the returning course instructor. The informed consent process required approximately 15 minutes.

The course instructor then distributed the Blue scenario packets that the researcher had previously provided. At the conclusion of the class session, the instructor collected all exercise materials.

#### *Second class session*

The course instructor distributed the Gold scenario packets. After students completed the Gold scenario, they then completed the post-study Participant Study Experience Survey. At the conclusion of the class session, the course instructor collected all exercise materials.

Approximately one week after the end of the exercise, the researcher collected copies of the scenario question documents and collected all other exercise materials from the course instructors. The ID anonymized consent forms were collected at semester end.

### **Study dataset development**

#### *Question scoring*

Each class session generated a set of completed scenario question response documents. Each question was scored horizontally across all question response documents before scoring the next question. A correct answer determination required a correct response to a Y/N or other binary choice, or an amount, and a written explanation that supported the response. Question answer determinations were guided by the Scenario Solutions document (Appendix E). The question solutions presented in the Scenario Solutions document are considerably more comprehensive and detailed compared to what would be expected from student work on an in-class exercise. Concise explanations were sufficient for students to demonstrate how applicable Tax Code provisions applied to scenario facts and circumstances.

## Data preparation

Several adjustments were applied to the initial dataset prior to and subsequent to scoring of the question response documents.

1. The researcher became aware, in conversation with the site 2 course instructor after the undergraduate participants had completed the Blue scenario, that collaboration is normally permitted in the course for in-class exercises, and that such collaboration did take place. Consequently, the researcher decided not to incorporate the site 2 undergraduate Blue scenario questions responses in the study dataset. For all the other experimental sessions, the course instructor announced that collaboration would not be permissible.

2. Three site 2 undergraduate scored Gold scenario response documents were deleted. Two of the three deletions were due to the complete absence, or nearly complete absence of written explanations in the response documents; the third deletion was one of a pair of student response documents that were essentially identical in all respects. In addition, one site 2 undergraduate student was absent. Hence, the total number of site 2 Gold scenario response documents =  $38 - (3 + 1) = 34$ .

3. Only one study participant correctly answered Blue scenario question 1. Consequently, this question was deleted from the study dataset, thereby reducing the number of Blue scenario question from 10 to nine.

Table 5 shows the participant composition and tuple volumes of the resultant study dataset. Scenario question responses were recorded in the tuple dataset structure described in Chapter 3.

Table 5

Composition of the Study Dataset

	Site 1 uGrad	Site 2 uGrad	Site 2 Grad	Questions /Scenario	Tuples
Blue scenario	25			9	225
Blue scenario		0			0
Blue scenario			16	9	144
Gold scenario	25			8	200
Gold scenario		34		8	272
Gold scenario			16	8	128
uGrad tuples					697
Grad tuples					272
Total tuples					969

### Experimental results

Aggregated undergraduate participant data (site 1 + site 2) and graduate participant data (site 2) are reported and analyzed separately. Table shows the percentage of undergraduate student participant correct answers by question ranked in order of percent correct answers, together with the number of Support cross-references and Scope cross-references associated with each question. Questions 2-10 (recall Question 1 was removed) are included in the Blue scenario, and questions 11-18 are included in the Gold scenario. (The Gold scenario questions were presented to participants numbered 1-8, but for analysis purposes, these questions are renumbered 11-18).

The ranked percent correct answers revealed a trimodal distribution where Group 1 includes questions with the lowest percentage of correct answers and Group 3 includes questions with the highest percentage of correct answers. Recall that hypothesis six predicts that reading Scope CR will have a greater negative impact on the number of correct answers to questions compared to the impact of reading Support CR. Table

shows that for all Group 1 questions, [Scope CR - Support CR] for all non-zero results is positive, whereas for all Group 3 questions the same quantity is negative.

Table 6.

## Undergraduate Participant Question Responses

Question	Correct Answers	Incorrect Answers	% Correct Answers		Scope CR	Support CR	[Scope - Support]
<b>Group 1</b>							
15	6	53	10.1		4	0	4
12	10	49	17.0		3	3	0
17	10	49	17.0		3	2	1
10	5	20	20.0		4	2	2
13	12	47	20.0		8	3	5
Totals	43	218	16.5	$\mu$	4.4	2.0	
				$\sigma$	2.1	1.2	
<b>Group 2</b>							
14	23	36	39.0		1	4	-3
5	10	15	40.0		4	2	2
9	10	15	40.0		4	1	2
7	11	14	44.0		1	0	1
18	26	33	44.0		1	3	-2
16	28	31	47.5		5	8	-3
8	12	13	48.0		4	1	3
Totals	120	157	43.3	$\mu$	2.9	2.7	
				$\sigma$	1.8	2.7	
<b>Group 3</b>							
11	49	10	83.0		2	4	-2
6	21	4	84.0		0	1	-1
3	22	3	88.0		2	3	-1
4	22	3	88.0		1	3	-2
2	24	1	96.0		2	6	-4
Totals	138	21	86.8	$\mu$	1.4	3.4	
				$\sigma$	0.9	1.8	

The same question distribution information for graduate student participants is shown in Table 7. Compared to the undergraduate question distribution, question 14 is now in group 1 and question 15 is now in group 2. With the exception of question 14, for

all Group 1 questions, [Scope CR - Support CR] for all non-zero results is positive, whereas for all Group 3 questions the same quantity is negative.

Table 7

Graduate student questions responses

Question	Correct Answers	Incorrect Answers	% Correct Answers		Scope CR	Support CR	[Scope - Support]
Grp. 1							
10	0	16	0		4	2	2
17	0	16	0		3	2	1
12	1	15	6.2		3	3	0
13	1	15	6.2		8	3	5
14	2	14	12.5		1	4	-3
Totals	4	76	5.0	$\mu$	3.8	2.8	
				$\sigma$	2.6	0.84	
Grp. 2							
15	4	12	25.0		4	0	4
5	5	11	31.2		4	2	2
18	5	11	31.2		1	3	-2
16	6	10	37.5		5	8	-3
8	7	9	43.8		4	1	3
7	8	8	50.0		1	0	1
9	8	8	50.0		4	1	3
Totals	43	69	38.4	$\mu$	3.3	2.1	
				$\sigma$	1.6	2.8	
Grp. 3							
11	11	5	68.8		2	4	-2
6	12	4	75.0		0	1	-1
3	13	3	81.2		2	3	-1
4	15	1	93.8		1	3	-2
2	16	0	100.0		2	6	-4
Totals	67	13	83.8	$\mu$	1.4	3.4	
				$\sigma$	0.9	1.8	



Figure 4 is a visualization of the Table 2 and Table 3 percent correct answers. The resolution of three question groups is better defined for the undergraduate participant data compared to the graduate participant data.

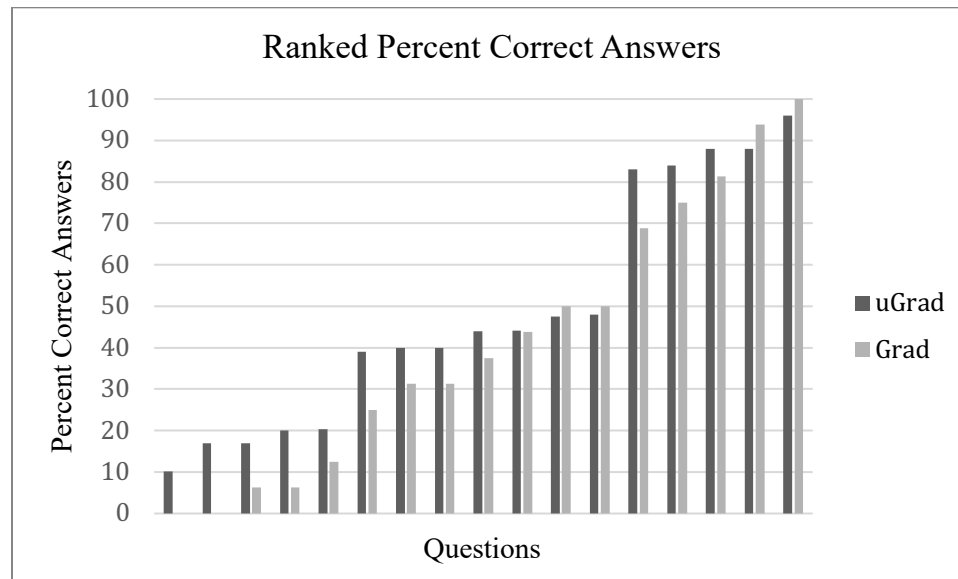


Figure 4. Trimodal distribution of percentage of correct question answers for undergraduate and graduate student participants based on Table and Table data. The two lowest percentage correct answers for graduate student participants are zero percent and are not displayed.

The [Scope – Support] values displayed in Table and Table show that greater numbers of Scope cross-references relative to numbers of Support cross-references are associated with questions that received low percentages of correct answers. Conversely, greater numbers of Support cross-references relative to numbers of Scope cross-references are associated with questions that received high percentages of correct answers. These comparisons provide suggestive evidence that Scope cross-references do have a negative impact on readability, and consequently here on task performance. The next section formally analyzes the study data in the context of the six study hypotheses.

## Hypothesis testing

### *Approach*

Five independent variables and six hypotheses were stated in Chapter 3. The five independent variables are: Tokens, Nodes, Tot\_CR (Total Cross-References), Scope Cross-References (Scope, or Scope CR), and Support Cross-References (Support, or Support CR).

Normalization was applied to these three variables to capture average cross-reference density, or frequency of reading interruptions. Densities calculated using Tokens and Nodes are displayed with .Tokens and .Nodes suffixes respectively.

Hypotheses 1 through 5 are tested using Binary Logistic Regression with one independent variable. Hypothesis 6 is tested using Binary Logistic Regression with two independent variables.

### *Multicollinearity*

Multicollinearity will be present in the dataset due principally to the following:

The use of Tokens and Nodes as independent variables. In the hierarchical organization of the Tax Code, provisions are represented as nodes, which in turn are collections of tokens that represent the Tax Code text. The sum of the tokens included in all nodes that participants were instructed to read = total tokens that participants were instructed to read. Hence, tokens and nodes represent the same information. Here, the effects of multicollinearity will be eliminated by conducting data analysis in separate but parallel processes, one for Tokens and .Token densities, and the other for Nodes and .Node densities.

The use of Total\_CR, Scope CR, and Support CR as independent variables. The study classified each cross-reference as type Scope or type Support, where  $\text{Total\_CR} = \text{Scope CR} + \text{Support CR}$ . Here, the effects of multicollinearity will be eliminated in data analysis by avoiding the use of Total\_CR together with one or both of Support CR and Scope CR.

#### Leverage points

In regression modeling, a leverage point is an extreme value of an independent, or predictor, variable along the x-axis where extreme means greater than a statistical package software calculated threshold value based in part on the average of all other values of the independent variable. Leverage points may bias the results of the regression model. A leverage point is potentially influential; i.e., removal of the observation(s) that created the leverage point may change aspects of the regression model such as the coefficient value, p-value of the variable, and the odds ratio. In Minitab, leverage points are reported as unusual values.

For the present study, the values of the independent variables were precalculated. The precalculated values were counts of non-normalized variables, or density values resulting from normalization of variables. For some models, the values of the independent variables were more evenly distributed in the X direction compared to other models. For example, Figure 5 is the dotplot of the precalculated values for the Scope.Tokens model. Minitab reported no unusual X values for the Scope.Tokens model, as the distance between the highest value and the average value of all other points was not so distant as to determine that value 0.962 constituted a leverage point.



Figure 5. Distribution of Scope.Tokens values. The scale is from 0.0 to 1.0. The highest value is 0.962. The two largest gaps are between 0.415 and 0.613 (gap length = 0.198), and between 0.613 and 0.870 (gap length = 0.257).

Figure 5 is the dotplot of the precalculated values for the Tot\_CR.Nodes model. In this case, Minitab did report unusual X values, as the distance between the value 2.0 and the average of the other values was sufficiently extreme to identify unusual X values. In this case, Blue Scenario Q10 was the source of the precalculated Tot\_CR.Nodes value = 2.0. Minitab labeled all 25 of the undergraduate responses, and all 16 of the graduate responses, to Q10 as unusual X values. In this study, if an independent variable is considered to be a leverage point, then the general case is that the group of all questions responses (observations) attached to the source scenario question may be influential and are candidates for removal from the model.



Figure 6. Distribution of the Tot\_CR.Nodes values. The scale is from 0.0 (not shown) to 2.0. The lowest value is 0.286. There is a gap between 1.0 and 2.0 (gap length = 1.0).

Since in the presence of a leverage point, the numbers of potential influential observations will be large, and typically larger than anticipated for the concept of influential points, leverage points and potential influential points were not considered when evaluating the results of the logistic regressions to test the study hypotheses. This matter is revisited in the following chapter. The complete set of independent variable dotplots is available in Appendix G.

### *Hypothesis testing design*

The magnitude of the values of the .Node densities are on average 25 times greater compared to the magnitude of the .Token densities. To provide for more meaningful interpretation, the .Token densities are scaled up to the same magnitude of the .Node densities by multiplying the calculated .Token densities by a factor of 25.

Hypothesis 1 is tested using the non-normalized Tokens independent variable, and Hypothesis 2 is tested using the non-normalized Nodes independent variable. Hypotheses three through five are each tested with three models: a model using one independent variable and two models using .Tokens and .Nodes normalizations of that independent variable. Hypothesis 3 is tested using Tot\_CR. Hypothesis 4 and Hypothesis 5 are designed for more granular analysis and insight, and so separately test the two components of Tot\_CR, Scope CR and Support CR. Hypothesis 6 is designed to compare the relative effects of participants reading Scope CR and reading Support CR on their development of question answers. Undergraduate participant data and graduate participant data are analyzed separately.

The presentation of each hypothesis testing model is as follows.

1. The hypothesis is stated
2. The model name(s) is stated.
3. The results of applying the model to the tuple dataset are displayed with the following abbreviations: Coeff (Coefficient); OR (Odds Ratio); CI (Confidence Interval). Effect is calculated as  $(OR - 1)$  and is expressed as a percent.
4. The analyses of the data models are presented. In cases where the analysis shows agreement between the results of the undergraduate participant model and

the graduate participant model, the results are discussed in the same paragraph. In such cases, for written conciseness, if the p-values associated with the models differ, the convention is to state the p-value associated with the undergraduate participant model followed by the p-value for the graduate participant model.

5. For Hypothesis 1 and Hypothesis 2, a conclusion is stated about whether the data supported or did not support the hypothesis.

6. For Hypothesis 3 through Hypothesis 6, an overall conclusion follows the individual model results regarding whether the data supported or did not support the hypothesis.

**H1:** The number of tokens a participant is required to read while formulating a question response is predictive of whether the answer is correct or incorrect.

Model: Tokens

Table 8

H1 model results

Undergraduate participants

<b>Coeff</b>	<b>p-value</b>	<b>R<sup>2</sup></b>	<b>AIC</b>	<b>OR</b>	<b>Effect</b>	<b>95% CI for OR</b>
-0.000038	0.9439	0.00%	957	0.9999	-0.01%	(0.999, 1.001)

Graduate participants

-0.000037	0.9679	0.00%	374	0.9999	-0.01%	(0.998, 1.002)
-----------	--------	-------	-----	--------	--------	----------------

For both participant groups, there is no statistically significant relationship between the number of tokens read by participants while developing answers to questions and the probability of a correct answer ( $p = .94$ ;  $p = .97$ , and the 95% CI for OR includes the value 1 (equal odds)). The data do not support the hypothesis and so the hypothesis is rejected, and the null hypothesis is accepted: The number of tokens a participant is required to read while formulating a question response is not predictive of whether the

answer is correct or incorrect. This result is unexpected, and will be discussed in the following chapter.

**H2:** The number of nodes a participant is required to read while formulating a question response is predictive of whether the answer is correct or incorrect.

Model: Nodes

Table 9

H2 model results

Undergraduate participants

Coeff	p-value	R <sup>2</sup>	AIC	OR	Effect	95% CI for OR
0.00438	0.6852	0.02%	957	1.00439	0.439%	(0.983, 1.026)

Graduate participants

0.00894	0.6103	0.07%	374	1.00898	0.898%	(0.975, 1.044)
---------	--------	-------	-----	---------	--------	----------------

For both participant groups, there is no statistically significant relationship between the number of nodes read by participants while developing answers to questions and the probability of a correct answer ( $p = .69$ ;  $p = .61$ , and the 95% CI for OR includes the value 1 (equal odds)). The data do not support the hypothesis and the hypothesis is rejected, and the null hypothesis is accepted: The number of nodes a participant is required to read while formulating a question response is not predictive of whether the answer is correct or incorrect. This result is unexpected, and will be discussed in the following chapter.

**H3:** The density of total cross-references a participant will encounter while reading required provisions is predictive of whether the answer is correct or incorrect.

H3 is tested with three models: Tot\_CR, Tot\_CR.Tokens; Tot\_CR.Nodes

Table 10

H3 model results

## Undergraduate participants

Variable	Coeff	p-value	R <sup>2</sup>	AIC	OR	Effect	95% CI for OR
Tot_CR	-0.0340	0.1802	0.19%	955	0.9667	-3.33%	(0.920, 1.016)
Tot_CR.Tokens	-0.9764	<0.0001	1.75%	941	0.3767	-62.33%	(0.231, 0.614)
Tot_CR.Nodes	-0.7607	0.0019	1.11%	947	0.4673	-53.27%	(0.289, 0.755)

## Graduate participants

Variable	Coeff	p-value	R <sup>2</sup>	AIC	OR	Effect	95% CI for OR
Tot_CR	-0.0985	0.0289	1.36%	369	0.9062	-9.38%	(0.830, 0.990)
Tot_CR.Tokens	-2.2242	<0.0001	6.64%	349	0.1082	-89.18%	(0.039, 0.299)
Tot_CR.Nodes	-1.4579	0.0005	4.39%	358	0.2327	-76.73%	(0.103, 0.526)

## Model: Tot\_CR

For undergraduate participants, there is no statistically significant relationship between the total number of cross-references a participant will encounter while reading required provisions and the probability of a correct answer ( $p = .18$ , and the 95% CI for OR includes the value 1.0 (equal odds)).

For graduate participants, there is a statistically significant relationship between the total number of cross-references a participant will encounter while reading required provisions and the probability of a correct answer ( $p = .029$ ). The Tot\_CR coefficient is negative, and so the odds of a correct response will decrease, here by 9.4%, for each additional cross-reference read by graduate participants.

## Model: Tot\_CR.Tokens

For both participant groups, there is a statistically significant relationship between the Tokens density of total cross-references a participant will encounter while reading required provisions and the probability of a correct answer ( $p < .001$ ). The Tot\_CR.Tokens coefficient is negative, and so the odds of a correct response will



decrease, here by 62% and 89% for undergraduate and graduate participants respectively, for each additional cross-reference per 25 tokens read by participants.

Model: Tot\_CR.Nodes

For both participant groups, there is a statistically significant relationship between the Nodes density of total cross-references a participant will encounter while reading required provisions and the probability of a correct answer ( $p = .002$ ;  $p < .001$ ). The Tot\_CR.Nodes coefficient is negative, and so the odds of a correct response will decrease, here by 53% and 77% for undergraduate and graduate participants respectively, for each additional cross-reference per node read by study participants.

### H3 conclusion

Based on the statistically significant Tot\_CR.Tokens and Tot\_CR.Nodes models, Hypothesis 3 is accepted. There is a negative relationship between reading provisions that contain cross-references and the subsequent decreasing odds of a correct response.

**H4:** The density of Scope cross-references a participant will encounter while reading required provisions is predictive of whether the answer is correct or incorrect.

H4 is tested with three models: Scope; Scope.Tokens; Scope.Nodes

Table 11

H4 model results

Undergraduate participants

Variable	Coeff	p-value	R <sup>2</sup>	AIC	OR	Effect	95% CI for OR
Scope	-0.2980	<0.0001	5.47%	905	0.7423	-25.77%	(0.680, 0.810)
Scope.Tokens	-2.7328	<0.0001	7.92%	882	0.0650	-935.0%	(0.033, 0.129)
Scope.Nodes	-2.8526	<0.0001	7.60%	885	0.0577	-942.3%	(0.027, 0.125)

## Graduate participants

Variable	Coeff	p-value	R <sup>2</sup>	AIC	OR	Effect	95% CI for OR
Scope	-0.4063	<0.0001	8.25%	343	0.6661	-33.39%	(0.568, 0.781)
Scope.Tokens	-3.2351	<0.0001	10.71%	334	0.0394	-960.6%	(0.013, 0.123)
Scope.Nodes	-3.1918	<0.0001	10.01%	337	0.0411	-958.9%	(0.011, 0.148)

## Model: Scope

For both participant groups, there is a statistically significant relationship between the number of Scope cross-references a participant will encounter while reading required provisions and the probability of a correct answer ( $p < .001$ ). The Scope coefficients are negative, and so for undergraduate and graduate participants respectively, the odds of a correct response will decrease, here by 26% and 33%, for each additional Scope cross-reference read by participants.

## Model: Scope.Tokens

For both participant groups, there is a statistically significant relationship between the density of Scope cross-references a participant will encounter while reading required provisions and the probability of a correct answer ( $p < .001$ ). The Scope.Tokens coefficients are negative, and so for undergraduate and graduate students respectively, the odds of a correct response will decrease, here by 935% and 960%, for each additional Scope cross-reference per 25 tokens read by participants.

## Model: Scope.Nodes

For both participant groups, there is a statistically significant relationship between the density of Scope cross-references a participant will encounter while reading required provisions and the probability of a correct answer ( $p < .001$ ). The Scope.Nodes coefficients are negative, and so for undergraduate and graduate students respectively, the

odds of a correct response will decrease by 940% and 960% for each additional Scope cross-reference per node read by participants

#### H4 conclusion

Based on the statistically significant Scope, Scope.Tokens, and Scope.Nodes models, Hypothesis 4 is accepted. The nearly identical effects associated with the Scope.Tokens and Scope.Nodes models indicate a strong negative relationship between study participants reading provisions that contain Scope cross-references and the subsequent decreasing odds of developing a correct response.

Here, the surprisingly high 10X magnitude of the readability effects derived from the Odds Ratios of the Scope.Tokens and Scope.Nodes models is especially noteworthy and is revisited in in the following chapter.

**H5:** The density of Support cross-references a participant will encounter while reading required provisions is predictive of whether the answer is correct or incorrect.

Hypothesis 5 is tested with three models: Support; Support.Tokens; Support.Node

#### Table 12

##### H5 model results

##### Undergraduate participants

Variable	Coeff	p-value	R <sup>2</sup>	AIC	OR	Effect	95% CI for OR
Support	0.1842	<0.0001	2.51%	933	1.2022	20.22%	(1.11, 1.30)
Support.Tokens	1.7107	<0.0001	2.15%	937	5.5326	453.26%	(2.61, 11.7)
Support.Nodes	1.6559	<0.0001	2.54%	933	5.2377	423.77%	(2.66, 10.3)

##### Graduate participants

Variable	Coeff	p-value	R <sup>2</sup>	AIC	OR	Effect	95% CI for OR
Support	0.1120	0.0709	0.89%	371	1.1186	11.86%	(0.991, 1.26)
Support.Tokens	0.4020	0.5108	0.12%	373	1.4948	49.48%	(0.451, 4.95)
Support.Nodes	0.3824	0.4246	0.17%	373	1.4658	46.58%	(0.573, 3.75)

#### Model: Support

For undergraduate participants, there is a statistically significant relationship between the total number of Support cross-references a participant will encounter while reading required provisions and the probability of a correct answer ( $p < .001$ ). The Scope coefficient is positive, and so the odds of a correct response will increase, here by 20%, for each additional Support cross-reference read by undergraduate participants.

For graduate participants, there is no statistically significant relationship between the total number of Scope cross-references a participant will encounter while reading required provisions and the probability of a correct answer ( $p = .071$ , and the 95% CI for OR includes the value 1.0 (equal odds)).

#### Model: Support.Tokens

For undergraduate participants, there is a statistically significant relationship between the Tokens density of Support cross-references a participant will encounter while reading required provisions and the probability of a correct answer ( $p < .001$ ). The Support.Tokens coefficient is positive, and so the odds of a correct response will increase, here by 450%, for each additional Support cross-reference per 25 tokens read by undergraduate participants. The 95% CI for OR is relatively wide.

For graduate participants, there is no statistically significant relationship between the density of Support cross-references a participant will encounter while reading required provisions and the probability of a correct answer ( $p = .51$ , and the 95% CI for OR includes the value 1.0 (equal odds)).

Model: Support.Nodes

For undergraduate participants, there is a statistically significant relationship between the Nodes density of Support cross-references a participant will encounter while reading required provisions and the probability of a correct answer ( $p < .001$ ). The Support.Nodes coefficient is positive, and so the odds of a correct response will increase, here by 425%, for each additional Support cross-reference per node read by undergraduate participants. The 95% CI for OR is relatively wide.

For graduate participants, there is no statistically significant relationship between the Nodes density of Support cross-references a participant will encounter while reading required provisions and the probability of a correct answer ( $p = .51$ , and the 95% CI for OR includes the value 1.0 (equal odds)).

#### H5 conclusion

Based on the statistically significant undergraduate Support, Support.Tokens, and Support.Nodes models, Hypothesis 5 is accepted. The magnitudes of the effects associated with the Support.Tokens and Support.Nodes models indicate a positive relationship between study participants reading provisions that contain Support cross-references and the subsequent increasing odds of developing a correct response. The three graduate participant models showed no statistically significant results. This result is not as expected and will be discussed in the next chapter.

**H6:** Questions that require participants to read provisions with a greater density of Scope cross-references will have a smaller number of correct responses compared to questions that require participants to read provisions with a greater density of Support cross-references.

The Hypothesis 4 analysis showed substantial negative effects on the probability of a correct answer due to study participants reading scope cross-references. Conversely, the Hypothesis 5 analysis showed substantial positive effects on the probability of a correct answer due to study participants reading support cross-references. Taken together, the H4 and H5 analyses support the acceptance of Hypothesis 6.

However, for consistency of approach, Hypothesis 6 is tested independently with three models each including two independent variables:

[Scope, Support]; [Scope.Tokens, Support.Tokens]; [Scope.Nodes, Support.Nodes]

Table 13

H6 model results

Undergraduate participants

Variable	Coeff	p-value	Adj R <sup>2</sup>	AIC	OR	Effect	95% CI for OR
Scope	-0.3374	<0.0001	8.88%	873	0.7136	-28.64%	(0.651, 0.782)
Support	0.2271	<0.0001			1.2549	25.49%	(1.162, 1.355)
Scope.Tokens	-2.5616	<0.0001	9.51%	867	0.0772	-922.8%	(0.041, 0.146)
Support.Tokens	1.7322	<0.0001			5.6532	465.32%	(2.470, 12.94)
Scope.Nodes	-2.5106	<0.0001	8.80%	873	0.0812	-918.8%	(0.039, 0.167)
Support.Nodes	1.4287	0.0003			4.1732	317.32%	(1.917, 9.084)

Graduate participants

Variable	Coeff	p-value	Adj R <sup>2</sup>	AIC	OR	Effect	95% CI for OR
Scope	-0.4280	<0.0001	9.33%	339	0.6518	-34.82%	(0.555, 0.766)
Support	0.1601	0.0139			1.1737	17.37%	(1.033, 1.333)
Scope.Tokens	-3.3374	<0.0001	10.24%	336	0.0355	-964.5%	(0.010, 0.122)
Support.Tokens	-0.3707	0.6073			0.6903	-30.97%	(0.168, 2.838)
Scope.Nodes	-3.2334	<0.0001	9.47%	339	0.0394	-960.6%	(0.001, 0.156)
Support.Nodes	-0.0989	0.8622			0.9058	-9.42%	(0.297, 2.767)

### H6 conclusion

For both participant groups, the coefficients of the Scope, Scope.Tokens, and Scope.Nodes model variables are all negative and statistically significant ( $p < .001$ ). For the undergraduate participants, the coefficients of the Support, Support.Tokens, and Support.Nodes model variables are all positive and statistically significant ( $p < .001$ ). For the graduate participants, the Support coefficient is positive and statistically significant ( $p = .014$ ) while the Support.Tokens and Support.Nodes coefficients are negative and not statistically significant ( $p = .61$  and  $p = .86$ , respectively). These results are consistent with the results reported for the Hypothesis 4 models and the Hypothesis 5 models. Since the negative Scope coefficient values have been shown in the Hypothesis 4 discussion to be associated with a decrease in the probability of a correct answer, and the positive Support coefficient values have been shown in the Hypothesis 5 discussion to be associated with an increased probability of a correct answer, Hypothesis 6 is accepted.

### **Post-test participant experience survey**

Immediately after completing the Gold scenario, participants completed the Likert-scaled Participant Study Experience Survey (Appendix H).

- (1) I thought the sections were easy to understand.
- (2) I thought the sections were too wordy.
- (3) I thought the sections were difficult to understand.
- (4) I thought the language used in the sections was too technical.
- (5) I thought the sections were clearly written.
- (6) I think it will be easy to remember the content of the sections tomorrow.
- (7) I am confident that all or most of my answers to the scenario questions are correct.

Figure 7 displays the undergraduate participant survey results, and Figure 8 displays the graduate participant survey results

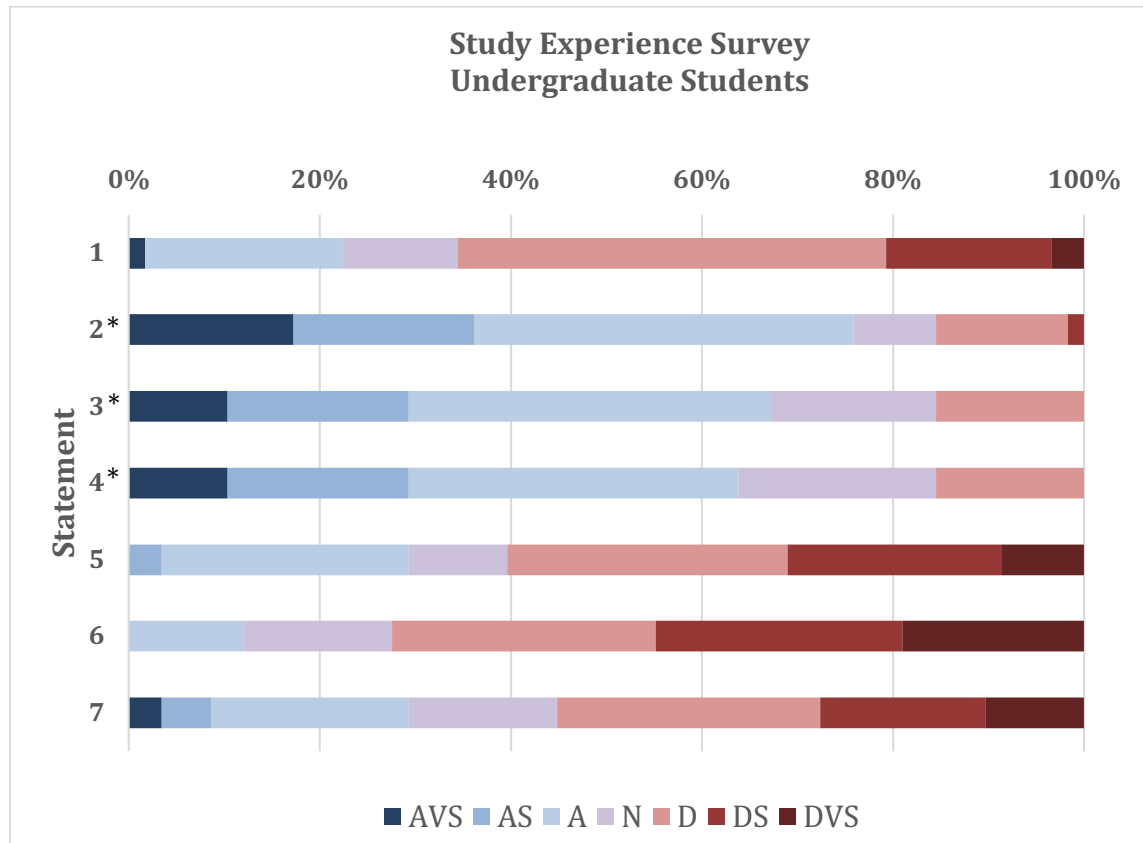


Figure 7. Undergraduate participant survey results. N=58. Likert-scale endpoints are AVS (Agree Very Strongly) and DVS (Disagree Very Strongly). Midpoint is N (Neither Agree or Disagree). \* Statements 2, 3, and 4 are expressed with negative polarity language; other statements are expressed with positive polarity language.

Compared to the undergraduate participants, graduate participants found the experimental task more challenging. Graduate participant midpoint (neutral) responses tended to be close to the Likert-scale endpoints, whereas the neutral responses for undergraduate participants tended to be more distant from the endpoints. The consequence of this neutral response location difference is that the graduate participant responses included more extreme Likert values compared to the undergraduate participant's relative absence of extreme Likert values.



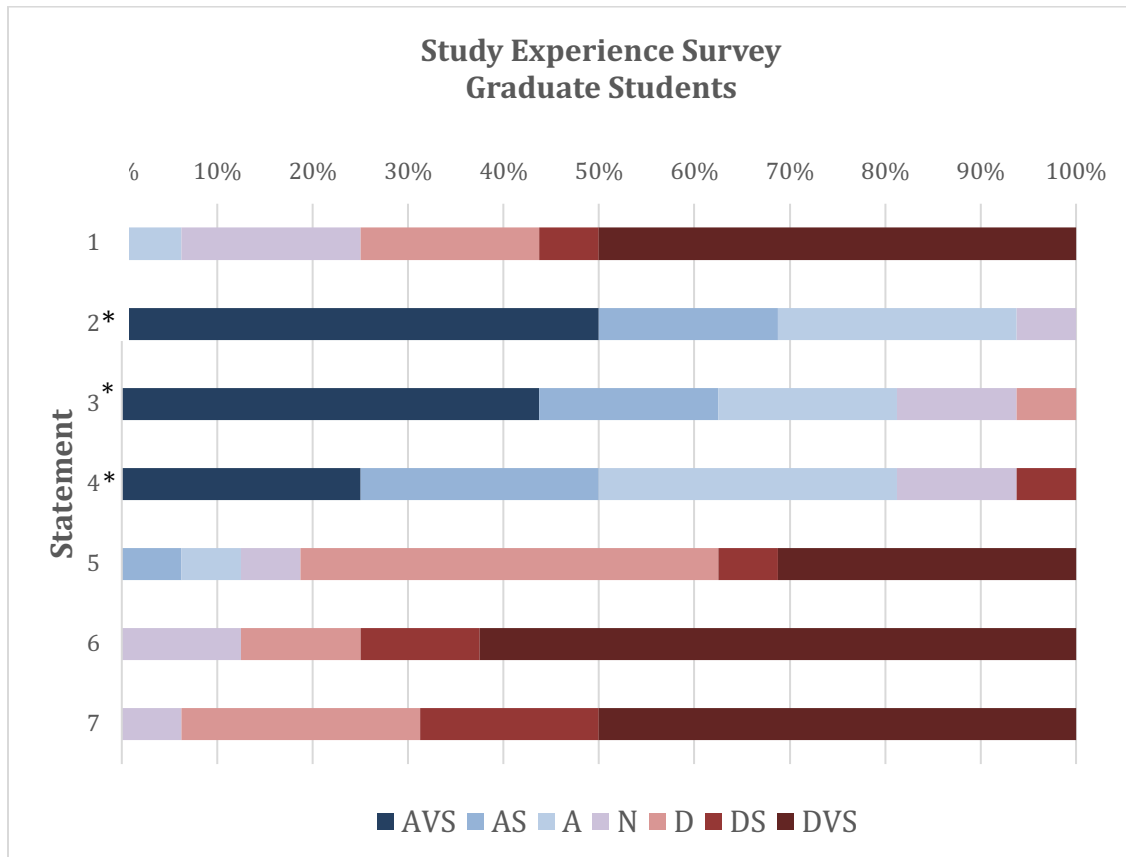


Figure 8. Graduate participant survey results. N=16. Likert-scale endpoints are AVS (Agree Very Strongly, and DVS (Disagree Very Strongly). Midpoint is N (Neither Agree or Disagree). \* Statements 2, 3, and 4 are expressed with negative polarity language; other statements are expressed with positive polarity language.

Graduate participant responses for statements 1, 2, and 3 were 50% (DVS), 50% (AVS), and 44% (AVS) respectively, and for statements 6, and 7 were 63%, and 50% respectively. In comparison, the largest extreme value for undergraduate participants was 19% DVS (statement 6).

With reference to the section readability statements, the developers of the survey instrument (Koch & Karlinsky, 1984) used statements one (sections easy to understand) and statement three (sections difficult to understand) to test for within subject consistency. For the present study, statements two through five are assessed for participant responses about section readability. The assessment shows directional

agreement between the two participant groups, and consistent with the above, graduate participants reported a higher level of readability difficulty compared to the undergraduate participants.

Considering statements two through four, the aggregate percentages for {AVS, AS, A} for graduate participants were 94%, 81%, and 81%, and for undergraduate participants, 76%, 67%, and 64%. For statement five, the aggregate percentages for {D, DS, DVS} were 81% and 60% for graduate and undergraduate participants respectively. To varying degrees, readability challenges are reflected in participant responses.

For statement six, section information retention, the aggregate percentages for {D, DS, DVS} were 81% and 60% for graduate and undergraduate participants respectively. Finally, for statement seven, confidence about correctness of question answers, the aggregate percentages for {D, DS, DVS} were 94% and 55% graduate and undergraduate participants respectively.

### **Delimitation**

The researcher obtained informed consent by visiting the classroom sites. The visit may have impacted the study participants attitudes and intentionality toward the in-class exercise. In addition, four absent participants requested and were granted make-up sessions. The informed consent anonymity protocol was followed, so it was not possible to delete the scenario question documents of these participants from the study dataset. Finally, the absence of the course instructor at one of the study sites for the second in-class exercise session and the presence of a substitute faculty proctor may have in some way changed the students sense of the classroom environment.

## Summary and conclusion

The objective of this study was to determine whether, and to what extent, reading two types of cross-references, Scope and Support, embedded in the text of the Tax Code are sources of reading complexity. Tokens and Nodes, two related measures of Tax Code text volume were included in hypothesis modeling. An initial descriptive analysis provided a first evidence that reading Scope type cross-references has a substantial negative impact on participants task performance. This analysis was followed by formal Binary Logistic Regression modeling. The key findings of the formal analysis are as follows:

- The number of tokens a participant was required to read while formulating a question response was not predictive of whether the answer is correct or incorrect. This is an unexpected result and is discussed in the following chapter.
- The number of nodes a participant was required to read while formulating a question response was not predictive of whether the answer is correct or incorrect. This is an unexpected result and is discussed in the following chapter.
- The density of total cross-references (type Scope + type Support) a participant encountered while reading required provisions was predictive of whether the answer was correct or incorrect. Increased densities of total cross-references reduced the odds of a correct answer.
- The density of Scope cross-references encountered by participants while reading required provisions was predictive of whether the answer was correct or incorrect. Increased densities of Scope cross-references reduced the odds of a correct answer.

- The density of Support cross-references encountered by participants while reading required provisions was predictive of whether the answer was correct or incorrect. Increased densities of Support cross-references increased the odds of a correct answer. This is an unexpected result and is discussed in the following chapter.
- Questions that required participants to read provisions with a greater density of Scope cross-references had a smaller number of correct responses compared to questions that required participants to read provisions with a greater density of Support cross-references.
- To varying but substantial degrees, participants reported readability challenges associated with their reading of required scenario Tax Code sections.

#### Conclusion

Based on the data analysis reported in this chapter, reading type Scope cross-references embedded in the text of the Tax Code is a primary and substantial source of increased reading complexity.

On the other hand, reading type Support cross-references embedded in the text of the Tax Code is not a source of increased reading complexity, but rather is a source of reduced reading complexity.

## Chapter 5

### Conclusions, Implications, Recommendations, and Summary

#### Introduction

The presence of cross-references embedded in U.S. statutory and similarly written regulatory text are frequently highlighted as a major contributor to readability complexity. The study goal was to determine to what extent cross-references encountered while reading statutory text impair readability and thus introduce difficulties in correctly applying statutory provisions. The study objective was to collect and analyze experimental data from individuals presented with the task of reading specified sections of statutory text and then developing answers to questions based on these readings. The statutory sections included varying numbers and types of cross-references.

The U.S. Tax Code, formally identified as U.S.C. Title 26 – Internal Revenue Code, was selected as the source of statutory text. Over several decades, a vast and expanding literature has developed about various aspects of Tax Code complexity. Scholars and practitioners alike have pointed to the presence of cross-references embedded in the text of the Tax Code’s provisions as a major source of such readability difficulties. A human-subjects experimental study was conducted to address the following research question: To what extent do cross-references contribute to a reader’s intellectual difficulty of collecting and synthesizing information within and across sections of income tax law text? To explore this question, cross-references were first considered in an undifferentiated manner, and then assigned to one of two cross-reference

types based on legal purpose. The first cross-reference type, Legal Scope, is used in the expression of exceptions to Tax Code provisions, and in the expression of limitations on the application of provisions. The second cross-reference type, Support, is intended to facilitate Tax Code readability principally via the addition of scaffolding language, and also by providing the textual location of definitions and of legal authority.

This chapter presents and discusses the study's hypotheses, conclusions and their implications, notes limitations of the study, suggests recommendations for future work, and concludes with a summary.

### **Study hypotheses and conclusions**

#### *Overview*

The study identified five independent variables that were expected to represent sources of negative impacts on Tax Code readability. Each of the following independent variables are counts of Tax Code text features participants were directed to read for developing answers to questions expressed in a tax scenario format similar to tax scenarios found in tax accounting textbooks and other materials used in tax accounting coursework. **Tokens**, a count of contiguous alphanumeric and special characters terminated by a space, is a measure of Tax Code text volume. Text volume includes headings, text of Tax Code provisions, hierarchical numbering identifiers, and cross-reference expressions. **Nodes** is a count of discrete Tax Code provisions. **Tot\_CR** is a count of the total number of cross-references. **Scope** is a count of scope type cross-references and **Support** is a count of support type cross-references, where  $Tot\_CR = Scope + Support$ .

Several analytic studies focused on legislative text incorporated normalized variables. Average sentence length, for example, was used in a study of New Zealand income tax law readability (Saw & Sawyer, 2010), and also used in a more general readability study of U.S. legislative and regulatory text (e.g., Curtotti et al. 2015b). In their complexity study of the United States Code, Katz and Bommarito (2014) included a token rank measure, based on normalizing tokens by section (tokens/section).

The study also included normalized variables. The objective of normalizing was to capture the density of cross-references per number of tokens read, and the density of cross-references per provisions read. Each of the independent variables Tot\_CR, Scope, and Support were normalized separately by Tokens (.Tokens) and by Nodes (.Nodes). The resulting normalized variables were Tot\_CR.Tokens, Tot\_Cr.Nodes; Scope.Tokens, Scope.Nodes; and Support.Tokens, Support.Nodes.

Six hypotheses were developed to test the extent to which the independent variables and the normalized independent variables impacted participants ability to develop correct answers to scenario questions. The first two hypotheses assessed the separate impacts of numbers of tokens and numbers of provisions read, respectively. The third hypothesis assessed the impact of numbers of cross-references read. The fourth and fifth hypotheses assessed the separate impacts of numbers of Scope cross-references read and of numbers of Support cross-references read. The sixth hypotheses compared the impact of Scope cross-references read to the impact of Support cross-references read. A discussion of the results of each hypothesis follows.

### *Hypothesis 1*

H1, the first hypothesis, that the number of tokens a participant is required to read while formulating a question response is predictive of whether the answer is correct or incorrect, was not supported by the Tokens model used to test H1. This result was unexpected. Intuitively, a requirement to read and assimilate larger numbers of Tax Code rules and details, and to process increasing numbers of cross-references, would have been found to have a negative impact on readability.

In the legislative domain, and in the Tax Code domain in particular, the relationship between information volume, or length, and readability is an unsettled matter. There is, however, general agreement that volume considered in isolation is not an appropriate predictor of readability. In the part of their study that addressed the entirety of the U.S.C., Katz and Bommarito (2014) conclude that “When considering the process of assimilating the information contained within a given element [i.e., a provision] of the [United States] Code, the volume of words [i.e., tokens] is important but by no means the exclusive [linguistic] property driving its complexity” (p. 354). Cauble’s (2019) analysis, focused principally on the Tax Code, echoes a similar conclusion about length and readability. “Although the length of applicable law is not entirely irrelevant to evaluate the complexity inherent in law, ...” (p. 349). Miller (1993) adds “Thus, the length and detail of tax rules, *along with* [emphasis added] their interconnectedness, are directly related to their elaborative complexity” (p. 12).

The independent variable Tokens was included in the present study to serve as a candidate factor to normalize three other independent variables, Tot\_CR, Scope, and Support, and thus obtain participant’s average frequency of encountering cross-



references/tokens per question. As Tokens is simply a token count, Tokens provided no information about the average frequency with which a participant would encounter a cross-reference in the Tax Code sections read by participants. Hence, from a metrics perspective, Tokens likely masked the hypothesized negative effects of interrupted reading, and so the Tokens model showed no relationship between Tokens and the number of correct scenario question answers.

### *Hypothesis 2*

H2, the second hypothesis, that the number of nodes a participant is required to read while formulating a question response is predictive of whether the answer is correct or incorrect, was not supported. This result was unexpected. In the present study, nodes are provisions of the Tax Code. Intuitively, a requirement to read and assimilate larger numbers of provisions would have been found to have a negative impact on readability.

Similar to the independent variable Tokens, the independent variable Nodes was included in the present study to serve as a candidate factor to normalize three other independent variables, Tot\_CR, Scope, and Support, and thus obtain participant's average frequency of encountering cross-references/provision per question. As Nodes is simply a count of a Tax Code section's hierarchical organization, Nodes provided no information about the average frequency with which a participant would encounter cross-references/provision. Hence, from a metrics perspective, Nodes likely masked the hypothesized negative effects of interrupted reading on experimental task performance, and so the Nodes model showed no relationship between Nodes and the number of correct scenario question answers.

*Hypothesis 3*

H3, the third hypothesis, that the density of total cross-references a participant will encounter while reading provisions is predictive of whether the answer is correct or incorrect, was supported. Here, normalizations of the Tot\_CR independent variable likely captured the hypothesized effects of interrupted reading on experimental task performance. The Tot\_CR.Tokens model results and the Tot\_CR.Nodes model results both show a negative relationship of similar magnitude between reading cross-references and task performance.

The literature reporting results of human-subjects task performance studies that include a focus on legal cross-references is sparse. The finding of a negative relationship between reading cross-references in the Tax Code and task performance is supported by prior experimental work (Karlinsky & Koch, 1987). In their experimental study, Maxwell, Anton and Earp (2013) reported a negative relationship between reading cross-references in excerpts of regulatory and statutory text and correctly assigning these cross-references to a pre-determined cross-reference taxonomy.

*Hypothesis 4*

H4, the fourth hypothesis, that the density of scope cross-references a participant will encounter while reading provisions is predictive of whether the answer is correct or incorrect, was supported. Here, normalizations of the Scope independent variable likely captured the hypothesized effects of interrupted reading on task performance. The H4 Scope.Tokens model results and the H4 Scope.Nodes model results both show a strong negative relationship of nearly identical magnitude between reading cross-references and task performance.

Recall, the Scope type cross-reference was intended to capture the effects of reading cross-references that express exceptions or qualifications to provisions or to parts of provisions. In most of these cases, the cue phrase component of a cross-reference expresses the exception or qualification. Prior work focused on isolating the readability effects of exception language in legislative text appears sparse. However, from the earliest reports of Tax Code readability challenges and to the present time, cross-references that express exceptions have been a prominent commentary feature (e.g., Holmes, 1926; Hand, 1946; Bhatia, 1983; Rook, 1993; Coven, 2000; Oei & Osofsky, 2018).

#### *Hypothesis 5*

H5, the fifth hypothesis, that the density of Support cross-references a participant will encounter while reading provisions is predictive of whether the answer is correct or incorrect, is supported, but the directionality of the results was unexpected. For undergraduate participants, the H5 Support.Tokens model results, and the H5 Support.Nodes model results, both showed a positive, rather than the expected negative, relationship between reading Support cross-references and task performance. (The graduate participant results were not statistically significant). As normalizing the cross-reference independent variables in the H3 models and the H4 models likely captured the cross-reference readability effects, the same is considered to have been the case for the H5 models.

A possible explanation of the unexpected H5 model results is that other, more impactful effects, outweighed the reading interruption effect. First, Support cross-references, in general, do not contain troublesome exception language as do Scope cross-

references, and so may just simply be easier to read and follow. Second, recall that Bhatia's analysis noted that the draftsman uses, by design, what is termed in this study as Support-type cross-references to impart a measure of text-coherence in provisions to facilitate readability (Bhatia, 1983; 1987). Consequently, the text-cohering effect may have been of sufficient magnitude to have outweighed the reading interruption effect.

#### *Hypothesis 6*

H6, the sixth hypothesis, that questions that require participants to read provisions with a greater density of Scope cross-references will have a smaller number of correct responses compared to questions that require participants to read provisions with a greater density of Support cross-references, was supported. This conclusion follows from considering together the H4 model results and the H5 model results, as well as the corroborating H6 model results.

The H6 result finding is that cross-reference type, or function, will determine the relative degree of resultant readability effects. The foundational framework that underlies the Katz and Bommarito (2014) U.S.C. complexity analysis includes the concept of a cognitive processing cost related to reading and understanding legal provisions (p. 352). By extension, the finding in this study of differences in the readability effects of Scope cross-references compared to the readability effects of Support cross-references may very well be explained by differential cognitive processing costs, relatively high for Scope cross-references, and relatively low for Support cross-references.

## Conclusions

Within the scale of the study, the data analysis supports the following conclusions.

- A user's ability to collect, synthesize, and correctly apply statutory provisions will not be adversely affected by the volume of statutory text read or by the number of statutory provisions read.
- The presence of Support cross-references embedded in the statutory text will tend to improve readability and application accuracy.
- The presence of Scope cross-references embedded in the statutory text will strongly degrade readability and application accuracy.

## Implications

The study results are consistent with the nearly 100 years of commentary regarding the contributions of [Scope] cross-references to difficulties in reading, understanding, and applying provisions of the Tax Code. Given the apparent immortality of Tax Code reading complexity, a practical implication of this study is to focus more instruction on how to read the Tax Code, with some special emphasis on how to read cross-references. At least one textbook is available for such an instructional purpose (Gershon & Maine, 2019). Keeping in mind that the U.S.C. is a compilation of hundreds of federal statutes all presented in a standardized hierarchical format with a consistent alphanumeric naming scheme that enables the use of cross-references, instruction about how to effectively read the Tax Code, including how to reason about cross-references, should generalize to instruction for reading any statute included in the U.S.C.

A more strategic implication of this study is the need to begin to address the sources of Tax Code readability complexity in specific and also the sources of statutory

readability complexity in general. Here, the focus will be on the statutory drafting practices as they relate to the use and expression of cross-references. The question at hand is how to effectuate improved statutory readability.

As a practical matter, and certainly for the near-term, there are at least two strong, interrelated constraints present in the writing and presentation of the Tax Code that will tend to perpetuate the present end user Tax Code readability experience as described in Chapter 1. The first constraint is the hierarchical/network structure of the United States Code. Since 1925/1926, this organizing structure for statutory text has been in continual use as the presentation format for U.S. federal statutes (“Office of the Law Revision Council,” (n.d.)). The properties of this structure, namely hierarchical format together with an associated hierarchical numbering scheme, topical and sub-topical arrangement, and referential syntactical structures (i.e., cross-references) all combine to facilitate the use of cross-references as a relatively low-cost drafting practice for adding exceptions and qualifications to existing provisions, as well as integrating new provisions to a statute (Stevenson, 2014, pp. 1146-1149).

The second constraint is drafting practice. Coven (2000) concluded that quality improvements in the design and drafting of Tax Code provisions are long overdue, and argues with focused conviction that Tax Code readability difficulties are due, in large measure, to “bad [provision] design and drafting” (p. 64). Coven identifies seemingly haphazard use of cross-references as one of several sources of bad drafting practice together with examples of cross-references that are “unnecessary, wrong, or accidents waiting to happen” (the latter referring to the risk(s) of referencing provisions that are not foundational (permanent) but rather are more likely to be subject to change (p. 60).

While drafting improvements may indeed be overdue, they will not likely be forthcoming as custom and practice has imposed a measure of rigidity in the Tax Code drafting process. Oei and Osofsky (2019) interviewed 26 individuals who were or had been involved in the process of drafting U.S. tax legislation. The interview reports show that reliance on cross-references to help achieve a variety of Tax Code drafting goals is historically and deeply ingrained in practice, while the Tax Code in general is afflicted with “inertial tendencies” (p. 1341) that underlie a drafting process resistant to change. The following segment published 60 years ago as an editorial in a tax journal summarizes the end user Tax Code readability experiences of many over the publication history of Title 26, from its inclusion in the initial edition of the U.S.C in 1926 to the present time

The "General Rule," though dutifully stated at the outset, is immediately rendered valueless by an array of references and cross-references to limiting exceptions, asides and interpolations, so that only somewhere in the inner reaches of the section, buried under layers of subsections, paragraphs, subparagraphs and subsubparagraphs, if you can find it at all is the nub of it. (Complexity Compounded, 1959, p. 99).

Given a likely continuance of the U.S.C in its present format, and a similar continuance of Tax Code drafting practices, a path forward to eventually improved statutory readability may very well need to be initiated outside of the statutory drafting ecosystem. One approach is to develop a domain-specific statutory/regulatory readability measure. Such a measure could provide readability guidance to draftspersons which in turn could motivate changes in drafting practice and eventually result in a more readable Tax Code, and more generally more readable statutes.

There is persistent and intense interest in readability, as evidenced by a vast and expanding readability literature. DuBay (2004) provides a comprehensive history of the development of readability measurement. Initially (circa 1945), the motivation to

measure readability was to determine the suitability of textbooks for various grade levels. In more recent decades, and stimulated by the plain language movement, the domain of readability interest has expanded to include adult consumers of everyday legal documents such as bank loan contracts, rental agreements, insurance policies, medical consent forms, and privacy and data breach notices. The Flesch Reading Ease (FRE) Formula that ranks readability on a scale from 0 to 100 (higher is easier to read), and the accompanying Flesch-Kincaid Grade Level Formula that maps the Reading Ease measure to a minimum U.S. grade-level required for understanding what is being read, have achieved ubiquitous use among practitioners and researchers alike. However, these and similar readability formulas have not been validated for assessment of legal document readability and their use for such a purpose is periodically questioned along with raised concerns (Redish & Selzer, 1985; Karlinksy & Koch, 1987, p.23; Katz & Bommarito, 2014, note 12, p. 352; Curtotti et al. (2015b), p.59).

Advances in the fields of Natural Language Processing and Computational Linguistics provide present-day readability researchers with platforms and tools to conduct readability studies at increased scope and scale compared to the time periods when the classical readability formulas were developed (Crossley, Skalicky, & Dascalu, 2019). The availability of a readability measure validated for application in the statutory/regulatory domain has the potential for widespread awareness raising if not also for adoption by statutory draftpersons to highlight their work that are assessed to be at high levels of readability complexity.

Given the study results that highlight the presence of exception language as a major contributor to degraded readability, a future statutory readability measure will need



to explicitly incorporate cross-references. The number of Scope cross-references per text unit of interest can provide a reading interruption impact factor for inclusion in the measure. In addition, and importantly, the intended purpose, or semantics, of the Scope cross-references will also need to be included, since arguably, differences in semantics may require differing levels of cognitive processing and so have varying impacts on readability. Initial work on the automatic identification and semantic typing of cross-references has been reported (Sannier, Adedjouma, Sabetzadeh, & Briand, 2015, 2016; Sadeghian, et al., 2018).

### **Recommendations for future work**

#### *Replication of the study*

While the present study has demonstrated a method to assess the impact of cross-references on Tax Code readability, addressing the major study limitations will improve the scope and usefulness of future readability research in the Tax Code domain. The following identifies study limitations together with recommendations for addressing the limitations.

#### Participant population

The participant pool was limited to accounting students enrolled in required taxation courses. Their educational experiences in these courses may have been different across the two experimental sites. These students had no or minimal exposure to the actual Tax Code text. Although a brief written tutorial about the organization of the U.S.C. together with an annotated example of a Tax Code section fragment was provided to participants prior to their engagement with the scenario exercise, the tutorial may have

been insufficient for providing a working orientation to the Tax Code and for providing a comfortable level of familiarity to statutory text that includes cross-references.

Although the experimental task was a graded, in-class exercise, students were given the option, in the context of informed consent, to deny permission for their exercise work to be included in the study dataset. Four of the 83 students (5%) across the two sites denied permission.

In the planning stage of the study, the degree to which students would be intentional in their reading of required provisions and cross-references, and in their authoring of required explanations was noted as a study limitation. The degree of completeness of the required answer explanations varied considerably, and included instances of answers with no meaningful accompanying explanations. Two scenario question documents included no required explanations and were removed from the study data set.

Recommendation: extend the potential participant pool to graduate students pursuing a specialized degree in Taxation, and/or law school students with exposure to a taxation course. These populations will already be familiar with the organization of the Tax Code and with reading statutory language that includes cross-references, and will likely be more intentional in their approach to the experimental task.

#### Scenario scope and scale

Scenario design decisions included placing a reasonable limit the number of sections that participants would be instructed to read, and to exclude sections centrally focused on business topics. The motivation for these design decisions included factors such as constraints on classroom time available to conduct the study, and uniformly

matching experimental task complexity to the participant populations. (Undergraduate participants enrolled in introductory tax courses had no prior exposure to business topics, while graduate students were enrolled in a course that addressed business topics). The effect of this design decision was to limit the volume of text and number of provisions to be read, and to limit the number of different related topics participants would encounter, thereby constraining the number of cross-references participants would encounter. Constraining the number of cross-references also reduced the number of cross-reference chains, a likely readability complexity factor (Rook, 1993).

Recommendation: increase the scale of the experimental task by increasing the number of sections and provisions participants will need to read. Since in general each additional section will address a different topic, increasing the number of sections will tend to increase the volume of Tax Code provisions participants will read, will provide an opportunity to include more complex topics, and will tend to increase the number of cross-references participants will encounter.

#### Scenario question design

While study participants were directed to read specific provisions, to follow encountered cross-referenced provisions, and to provide brief written explanations of questions answers, the degree of adherence to reading requirements was generally not strongly determinable. In turn, the number of tokens, the number of provisions, and the number of cross-references participants actually read and considered was likewise not determinable.

Recommendation: Enhance the data collect method by incorporating sub-questions for each scenario question that require an explanation of the impact of each

cross-reference participants are required to read. This approach will increase the likelihood that all reading requirements will be undertaken and completed, and will provide opportunities for more fine-grained insights into relationships between reading cross-references and readability complexity. A second enhancement would be to create an instrumented, click recording online environment that contains the Tax Code provisions participants will be instructed to read. Measurements such as the order in which each provision and cross-reference were accessed, the number of times each provision and cross-reference were accessed, and the elapsed time between clicks will provide a rich data collection with potential to complement the insights obtainable from the use of sub-questions described above with additional fine-grained analysis.

#### Aggregation of cross-reference types

The study was designed to determine whether there are differential levels of readability complexity that result from reading type Scope cross-references compared to reading all other cross-references aggregated into type Support. The sheer number of defined terms in numerous sections of the Tax Code, together with the numbers of cross-references required in many cases to specify the text location of term definitions, is increasingly being identified as a likely source of readability complexity (Morrison, 2014; Benson, 2019).

Recommendation: disaggregate the study's Support type cross-reference so that cross-references that refer to term definitions are considered a separate type, or separate independent variable, and are modeled as such.

### Data analysis

The scenario design and scenario question development process did not take into account the distribution of independent variable values that resulted from the numbers of tokens, nodes, Scope cross-references, and Support cross-references that participants were required to read for developing answers to each scenario question. During the execution of logistic regression analysis, the statistical software package (Minitab) identified certain extreme values of independent variables as unusual values. The identification of unusual values was the result of non-uniform and sparse independent variable value distributions. Unusual values have the potential to substantially influence regression results.

Recommendation: achieve a more uniform distribution of independent variable values to minimize potential statistical issues associated with regression modeling. This will require close attention and consideration of Tax Code section selection, and scenario and question development, and may substantially increase the challenges of scenario and question design.

### *Investigation of exception language*

The study results highlighted very strong negative impacts of reading type Scope cross-references on Tax Code readability. Scope cross-references, which are present in other statutory texts, is a component of what can be termed the statutory exception language, that subset of the statutory language that expresses exceptions, broadly construed. The study included examples of how the semantics of Scope cross-reference cue phrases defined relationships among and between statutory provisions that required reasoning for technical problem solving. Cue phrases may contain, for example, language

elements such as conditionals (e.g., if, unless), connectives (e.g., and, or), and negations (e.g., notwithstanding, but for). These same exception language elements may also be present as well in statutory text that does not include cross-references. Given the potential impact effects of the statutory exception language on readability, experimental and other empirical work on the topic of reasoning about statutory provisions in the presence of exception language is arguably an attractive, high-payoff research area.

### **Summary**

A human-subjects experimental study was conducted to investigate to what extent do cross-references contribute to a user's intellectual difficulty of collecting and synthesizing information within and across sections of U.S. income tax law text. The major finding was that cross-references that participate in the expressions of exceptions to provisions of the income tax law have a strongly degrade Tax Code readability.

The study contributes in two ways to the large and expanding volumes of work that address income tax readability. First, to the knowledge of this researcher, the study is the first human-subjects experiment designed specifically to investigate the relationship between cross-references and readability complexity in the income tax domain. Second, the study preliminarily validates approximately 100 years of published rhetoric and more considered analysis regarding difficulties in reading, understanding, and applying the provisions of income tax law.

Improvements to this state-of-affairs are needed. The income tax law permeates and influences, on a sustained basis, the economic decisions and plans of individuals, families, for-profit businesses of all types, and tax-exempt organizations. The annual cost of tax law compliance costs in the U.S. is staggering, most recently estimated to require 8

billion hours of effort that represents an estimated economic value of \$365 billion (Brady, 2019). Among factors that contribute to tax law complexity and costs, the relative contribution of readability complexity is presently unknown and perhaps unknowable. However, achievement of enhanced readability would certainly be a positive development.

A feasible path forward to achieving enhanced income tax law readability, or more generally enhanced statutory and regulatory law readability, is the development of a validated method for measuring readability in the statutory/regulatory domain. The availability of such a method would likely receive a good measure of attention, and would have the potential to eventually improve drafting practices by highlighting, in an understandable way, the sources and impacts of readability difficulties. Along these lines, a complementary research effort focused on the statutory exception language will contribute to the development of a validated statutory readability measure.

## **Appendix A**

### **Participant Scenario Materials**

Organization and Naming of IRC Sections

Instructions and Notes

Blue Scenario

Blue Scenario Questions

Gold Scenario

Gold Scenario Questions



## The Organization and Naming of Internal Revenue Code Sections

All federal laws, including the IRC, are published using a standard format and order that is described in this reference note.

Sections contains several levels of subdivisions. The number of subdivisions and the types of subdivisions will vary from section to section. Subdivisions are labeled using upper and lower case alphabetic characters, integers, and roman numerals. The label is shown enclosed in a parenthesis.

Sections are numbered and named. The section number is preceded by the symbol §. The section name follows the section number.

Subdivisions appear in the order as shown in the table. Examples are based on an excerpt from §221, which is **Interest on education loans**. §221 is attached to this document.

Subdivision name	Label format	Examples from §221
subsection	Lower case alphabetic	(a) Allowance of deduction (b) Maximum deduction
paragraph	Integer	(1) In general (2) Limitation based on .....
subparagraph	Upper case alphabetic	(A) In general (B) Amount of reduction
clause	Lower case roman numeral	(i) (ii)
subclause	Upper case roman numeral	(I) (II)

The formal identifier of each subdivision is the collection of labels that start at the section level and lead to the subdivision of interest.

§221 subdivision	Identifier
Allowance of deduction	§221(a)
Maximum deduction	§221(b)
In general	§221(b)(1)
Limitation based on modified adjusted gross income	§221(b)(2)
In general	§221(b)(2)(A)
Amount of reduction	§221(b)(2)(B)
clause (i)	§221(b)(2)(B)(i)
Subclause (I)	§221(b)(2)(B)(i)(I)
Subclause (II)	§221(b)(2)(B)(i)(II)
clause (ii)	§221(b)(2)(B)(ii)

**§221. Interest on education loans****(a) Allowance of deduction**

In the case of an individual, there shall be allowed as a deduction for the taxable year an amount equal to the interest paid by the taxpayer during the taxable year on any qualified education loan.

**(b) Maximum deduction****(1) In general**

Except as provided in paragraph (2), the deduction allowed by subsection (a) for the taxable year shall not exceed \$2,500.

**(2) Limitation based on modified adjusted gross income****(A) In general**

The amount which would (but for this paragraph) be allowable as a deduction under this section shall be reduced (but not below zero) by the amount determined under subparagraph (B).

**(B) Amount of reduction**

The amount determined under this subparagraph is the amount which bears the same ratio to the amount which would be so taken into account as-

- (i) the excess of-
  - (I) the taxpayer's modified adjusted gross income for such taxable year, over
  - (II) \$50,000 (\$100,000 in the case of a joint return), bears to

- (ii) \$15,000 (\$30,000 in the case of a joint return).

## Instructions and Notes

### Instructions

The exercise asks you to read a tax scenario and to answer questions that are based on the facts and circumstances presented in the scenario. As reference material, you are provided with several sections of the Internal Revenue Code. These sections contain the actual text of the Internal Revenue Code as written by Congress.

The scenario is written and the questions are presented as if the year 2018 has already ended, and so you will be responding to questions as if year 2019 has already begun.

Each question will be answered by circling one of the Y / N choices (in one case, the choices are phrases), or by entering an amount in the space provided. In one question, the two choices are not Y / N. Each question requires a written explanation that you will enter in the textbox immediately following the question. The size of the textbox is not an indication of the expected length of your explanation. Try to write brief explanations. In your explanations, include the specific Tax Code provision or provisions you relied on when determining your answers.

The provided IRC sections are the authoritative source for answering questions. As guidance for using the IRC sections, each question contains specific IRC provisions you will need to read for developing your answer to each question. These specific provisions will be listed following the question, although in a few cases, there are also provisions included in the question itself that you will need to read as well. As you read certain provisions, you will encounter references to other provisions. You will need to read these referenced provisions. Occasionally, a listing of a question's provisions that you will need to read will include the phrase "but not". You should not read the sections following this phrase, as the provisions are not relevant to the scenario and the question being asked.

### Important notes

Some of the IRC sections that you will read contain references to provisions in other sections that are not being provided to you. Rather, these provisions are discussed below.

1. §170(b)(1)(A)(ii). This provision specifies the requirements for an organization to be deemed an *educational organization*. Whenever you encounter a reference to section 170(b)(1)(A)(ii), you can assume that the educational organization fully satisfies the requirements of §170(b)(1)(A)(ii).
2. §6013. This section contains the rules for married couples to file joint returns. You do not need to know these rules in order to answer certain scenario questions.

3. §151. This section provides the legal authority to allow the taxpayer to deduct the exemption amount from taxable income once a determination has been made that an individual qualifies as a dependent. If you make such a determination that an individual qualifies as a dependent, then consider the §151 requirements have been met.

§151(d). This provision is the exemption amount that the IRS periodically adjusts for changes in inflation. For 2018, this amount is \$4,150.

➡ Although the recent changes to the IRC set the exemption amount for dependents to be zero, there are other provisions IRC provisions that reference and use the exemption amount for other purposes.

4. §117(c)(2). This provision references certain kinds of educational programs that are not described in the IRC, but rather are described in other federal laws. In the context of the scenario, the names of the listed educational programs will be sufficient for your purposes.

#### Other items.

- (a). You may encounter reference to the phrase “this title”. “This title” means the IRC.
- (b). Ignore any references you encounter that refer to a chapter, a subchapter, or a part. These are references to the IRC table of contents.

## **Blue Scenario**

### Overview of the scenario

The participants in the scenario are Parents (Parent 1 and Parent 2), their child named Child, and Child's spouse named Spouse.

### The Parents

Parents, born in the U.S., are both aged 55, are married with one child named Child. Parents, while not divorced or legally separated, have lived apart during all of 2017 and 2018. Parents provide 100% of the cost of maintaining their respective households. Given these living circumstances, parents will file separate 2018 tax returns (and both will itemize deductions). Parent 1's AGI is \$60,000, and Parent 2's AGI is \$50,000. Parents live in the same city and work at different local universities.

### The Child and Child's Spouse

Child, age 21 in 2018, is married to Spouse. Child and Spouse live all year with Parent 1 in Parent 1's home. Spouse, age 21, is a full-time undergraduate student pursuing a degree at a local university, and earned \$6,000 in 2018. Spouse used the entire \$6,000 for spouse's own expenses. For personal reasons, Child and Spouse file separate returns.

### Matters related to Child's undergraduate degree program

Child is a full-time undergraduate student pursuing a degree at a local state university. In 2018, that university awarded Child a \$18,000 cash scholarship award and provided a \$2,000 stipend (payment) for undergraduate teaching assistant services that are required for undergraduate scholarship recipients. Child used the scholarship funds and the stipend for the following expenses: \$14,000 tuition; \$1,000 required fees; \$3,000 meals on campus; \$1,000 textbooks; \$1,000 required lab course supplies

### Child's support

Parents and Child agreed that for purposes of Child's support, Child would contribute, in addition to the scholarship funds and stipend, \$7,000 from personal savings, and that Parents would provide whatever support remained. Parent 1's tax advisor determined that Child's total 2018 support was \$39,000.

### Child's illness

During 2018, Child received medical care not fully reimbursed by health insurance. The type of medical care Child received is consistent with the Tax Code definition of what constitutes qualified medical expenses. The amount of out-of-pocket costs was \$12,000.

Parents agreed to equally share the out-of-pocket expenses. Accordingly, Parent 1 provided \$6,000 and Parent 2 provided \$6,000 to pay the out-of-pocket costs.













## Gold Scenario

### Overview of the scenario

The participants in the scenario are Parents (Parent 1 and Parent 2), their child named Child, and Child's spouse named Spouse.

### The Parents

Parents, born in the U.S., are both aged 55, are married with one child named Child. Parents, while not divorced or legally separated, have lived apart during all of 2017 and 2018. Parents provide 100% of the cost of maintaining their respective households. Given these living circumstances, parents will file separate 2018 tax returns (and both will itemize deductions). Parent 1's AGI is \$65,000, and Parent 2's AGI is \$55,000.

### The Child and Child's Spouse

Child, age 25 in 2018, is married to Spouse, also age 25. Child and Spouse live all year with Parent 1 in Parent 1's home. In 2018, Child is a full-time graduate student enrolled in an MBA degree program offered by a local university. During 2018, child engaged in uncompensated, volunteer activities. Spouse, a full-time graduate student enrolled in an MS degree program, earned \$6,000 working part-time in 2018, and used the entire \$6,000 for Spouse's own expenses. For personal reasons, Child and Spouse file separate returns.

### Matters related to Child's graduate degree program

Child's graduate program is offered by the University where Parent 1 is employed. The University offers tuition reductions (also known as tuition waivers) to their employees. Tuition reduction is an example of a type of employee fringe benefit termed a no-additional-cost (to the employer) service. In certain cases, the tuition reduction benefit can be used by individuals related to the employee. Even in such cases, the tuition reduction is considered to be the employee's benefit, and any tax liability that may be associated with the tuition reduction is the responsibility of the employee.

### Child's support

For 2018, Parent 1 provided 15% of Child's total support. Parent 2 contributed 10%, one of Child's grandparents contributed 35%, and Parent 1's brother contributed 20%. Taken together, these four individuals contributed a total of 80% for Child's support. For tax purposes, this type of support funding arrangement is known as a multiple support agreement. Child will use personal savings to cover the remaining 20% of total support. Parent 1's tax advisor determined that Child's total 2018 support was \$40,000.

(scenario continued on next page)

Child's illness

During 2018, Child became ill for three weeks and was not capable of self-care. Consequently, household and care services for Child were needed in Parent 1's home during the daytime hours while Parent 1 worked and Spouse attended classes and worked part-time. These services enabled Parent 1 to continue to go to work and so are deemed by the Tax Code to be employment-related expenses. Such expenses, subject to conditions and limitations, are allowable as a tax credit to the individual who paid for the employment-related expenses. The amount expended for these services was \$3,200, and Parents agreed to share the amount equally, each providing \$1,600. Separate from the employment-related expenses, all Child's medical expenses were reimbursed in full.













## Appendix B

### Scenario Tax Code Sections

- §2. Definitions and special rules
- §21. Expenses for household and dependent care services necessary for gainful employment
- §117. Qualified scholarships
- §132. Certain fringe benefits
- §152. Dependent defined
- §213. Medical, Dental, etc., expenses
- §7703. Determination of marital status

#### **Cross-Reference Category Legend** (not present on participant exercise materials)

Legal Authority

Terminological Explanation

Textual Mapping

Legal Scope

Stipulated Provision

## §2. Definitions and special rules

### (a) Definition of surviving spouse

#### (1) In general

For purposes of section 1, the term "surviving spouse" means a taxpayer-

- (A) whose spouse died during either of his two taxable years immediately preceding the taxable year, and
- (B) who maintains as his home a household which constitutes for the taxable year the principal place of abode (as a member of such household) of a dependent (i) who (within the meaning of section 152, determined without regard to subsections (b)(1), (b)(2), and (d)(1)(B) thereof) is a son, stepson, daughter, or stepdaughter of the taxpayer, and (ii) with respect to whom the taxpayer is entitled to a deduction for the taxable year under section 151.

For purposes of this paragraph, an individual shall be considered as maintaining a household only if over half of the cost of maintaining the household during the taxable year is furnished by such individual.

#### (2) Limitations

Notwithstanding paragraph (1), for purposes of section 1 a taxpayer shall not be considered to be a surviving spouse-

- (A) if the taxpayer has remarried at any time before the close of the taxable year, or
- (B) unless, for the taxpayer's taxable year during which his spouse died, a joint return could have been made under the provisions of section 6013 (without regard to subsection (a)(3) thereof).

#### (3) Special rule where deceased spouse was in missing status

If an individual was in a missing status (within the meaning of section 6013(f)(3)) as a result of service in a combat zone (as determined for purposes of section 112) and if such individual remains in such status until the date referred to in subparagraph (A) or (B), then, for purposes of paragraph (1)(A), the date on which such individual died shall be treated as the earlier of the date determined under subparagraph (A) or the date determined under subparagraph (B):

- (A) the date on which the determination is made under section 556 of title 37 of the United States Code or under section 5566 of title 5 of such Code (whichever is applicable) that such individual died while in such missing status, or
- (B) except in the case of the combat zone designated for purposes of the Vietnam conflict, the date which is 2 years after the date designated under section 112 as the date of termination of combatant activities in that zone.

### (b) Definition of head of household

#### (1) In general

For purposes of this subtitle, an individual shall be considered a head of a household if, and only if, such individual is not married at the close of his taxable year, is not a surviving spouse (as defined in subsection (a)), and either-

(A) maintains as his home a household which constitutes for more than one-half of such taxable year the principal place of abode, as a member of such household, of-

(i) a qualifying child of the individual (as defined in section 152(c), determined without regard to section 152(e)), but not if such child-

- (I) is married at the close of the taxpayer's taxable year, and
- (II) is not a dependent of such individual by reason of section 152(b)(2) or 152(b)(3), or both, or

(ii) any other person who is a dependent of the taxpayer, if the taxpayer is entitled to a deduction for the taxable year for such person under section 151, or

(B) maintains a household which constitutes for such taxable year the principal place of abode of the father or mother of the taxpayer, if the taxpayer is entitled to a deduction for the taxable year for such father or mother under section 151.

For purposes of this paragraph, an individual shall be considered as maintaining a household only if over half of the cost of maintaining the household during the taxable year is furnished by such individual.

**(2) Determination of status**

For purposes of this subsection-

(A) an individual who is legally separated from his spouse under a decree of divorce or of separate maintenance shall not be considered as married;

(B) a taxpayer shall be considered as not married at the close of his taxable year if at any time during the taxable year his spouse is a nonresident alien; and

(C) a taxpayer shall be considered as married at the close of his taxable year if his spouse (other than a spouse described in subparagraph (B)) died during the taxable year.

**(3) Limitations**

Notwithstanding paragraph (1), for purposes of this subtitle a taxpayer shall not be considered to be a head of a household-

(A) if at any time during the taxable year he is a nonresident alien; or

(B) by reason of an individual who would not be a dependent for the taxable year but for-

(i) subparagraph (H) of section 152(d)(2), or

(ii) paragraph (3) of section 152(d).

**(c) Certain married individuals living apart**

For purposes of this part, an individual shall be treated as not married at the close of the taxable year if such individual is so treated under the provisions of section 7703(b).

**(d) Nonresident aliens**

In the case of a nonresident alien individual, the taxes imposed by sections 1 and 55 shall apply only as provided by section 871 or 877.

**(e) Cross reference**

For definition of taxable income, see section 63

## §21. Expenses for household and dependent care services necessary for gainful employment

### (a) Allowance of credit

#### (1) In general

In the case of an individual for which there are 1 or more qualifying individuals (as defined in subsection (b)(1)) with respect to such individual, there shall be allowed as a credit against the tax imposed by this chapter for the taxable year an amount equal to the applicable percentage of the employment-related expenses (as defined in subsection (b)(2)) paid by such individual during the taxable year.

#### (2) Applicable percentage defined

For purposes of paragraph (1), the term "applicable percentage" means 35 percent reduced (but not below 20 percent) by 1 percentage point for each \$2,000 (or fraction thereof) by which the taxpayer's adjusted gross income for the taxable year exceeds \$15,000.

### (b) Definitions of qualifying individual and employment-related expenses

For purposes of this section-

#### (1) Qualifying individual

The term "qualifying individual" means-

- (A) a dependent of the taxpayer (as defined in section 152(a)(1)) who has not attained age 13,
- (B) a dependent of the taxpayer (as defined in section 152, determined without regard to subsections (b)(1), (b)(2), and (d)(1)(B)) who is physically or mentally incapable of caring for himself or herself and who has the same principal place of abode as the taxpayer for more than one-half of such taxable year, or
- (C) the spouse of the taxpayer, if the spouse is physically or mentally incapable of caring for himself or herself and who has the same principal place of abode as the taxpayer for more than one-half of such taxable year.

#### (2) Employment-related expenses

##### (A) In general

The term "employment-related expenses" means amounts paid for the following expenses, but only if such expenses are incurred to enable the taxpayer to be gainfully employed for any period for which there are 1 or more qualifying individuals with respect to the taxpayer:

- (i) expenses for household services, and
- (ii) expenses for the care of a qualifying individual.

Such term shall not include any amount paid for services outside the taxpayer's household at a camp where the qualifying individual stays overnight.

##### (B) Exception

Employment-related expenses described in subparagraph (A) which are incurred for services outside the taxpayer's household shall be taken into account only if incurred for the care of-

- (i) a qualifying individual described in paragraph (1)(A), or
- (ii) a qualifying individual (not described in paragraph (1)(A)) who regularly spends at least 8 hours each day in the taxpayer's household.

##### (C) Dependent care centers

Employment-related expenses described in subparagraph (A) which are incurred for services provided outside the taxpayer's household by a dependent care center (as defined in subparagraph (D)) shall be taken into account only if-

- (i) such center complies with all applicable laws and regulations of a State or unit of local government, and
- (ii) the requirements of subparagraph (B) are met.

##### (D) Dependent care center defined

For purposes of this paragraph, the term "dependent care center" means any facility which-

- (i) provides care for more than six individuals (other than individuals who reside at the facility), and
- (ii) receives a fee, payment, or grant for providing services for any of the individuals (regardless of whether such facility is operated for profit).

**(c) Dollar limit on amount creditable**

The amount of the employment-related expenses incurred during any taxable year which may be taken into account **under subsection (a) shall not exceed-**

- (1) \$3,000 if there is 1 qualifying individual with respect to the taxpayer for such taxable year, or
- (2) \$6,000 if there are 2 or more qualifying individuals with respect to the taxpayer for such taxable year.

The amount determined under paragraph (1) or (2) (whichever is applicable) shall be reduced by the aggregate amount excludable from gross income under section 129 for the taxable year.

**(d) Earned income limitation****(1) In general**

Except as otherwise provided in this subsection, the amount of the employment-related expenses incurred during any taxable year which may be taken into account under subsection (a) shall not exceed-

- (A) in the case of an individual who is not married at the close of such year, such individual's earned income for such year, or
- (B) in the case of an individual who is married at the close of such year, the lesser of such individual's earned income or the earned income of his spouse for such year.

**(2) Special rule for spouse who is a student or incapable of caring for himself**

In the case of a spouse who is a student or a qualifying individual described in subsection (b)(1)(C), for purposes of paragraph (1), such spouse shall be deemed for each month during which such spouse is a full-time student at an educational institution, or is such a qualifying individual, to be gainfully employed and to have earned income of not less than-

- (A) \$250 if subsection (c)(1) applies for the taxable year, or
- (B) \$500 if subsection (c)(2) applies for the taxable year.

In the case of any husband and wife, this paragraph shall apply with respect to only one spouse for any one month.

**(e) Special rules**

For purposes of this section-

**(1) Place of abode**

An individual shall not be treated as having the same principal place of abode of the taxpayer if at any time during the taxable year of the taxpayer the relationship between the individual and the taxpayer is in violation of local law.

**(2) Married couples must file joint return**

If the taxpayer is married at the close of the taxable year, the credit shall be allowed under subsection (a) only if the taxpayer and his spouse file a joint return for the taxable year.

**(3) Marital status**

An individual legally separated from his spouse under a decree of divorce or of separate maintenance shall not be considered as married.

**(4) Certain married individuals living apart**

If-

- (A) an individual who is married and who files a separate return-
  - (i) maintains as his home a household which constitutes for more than one-half of the taxable year the principal place of abode of a qualifying individual, and
  - (ii) furnishes over half of the cost of maintaining such household during the taxable year, and

(B) during the last 6 months of such taxable year such individual's spouse is not a member of such household,

such individual shall not be considered as married.



**(5) Special dependency test in case of divorced parents, etc.**

if-

- (A) section 152(e) applies to any child with respect to any calendar year, and
- (B) such child is under the age of 13 or is physically or mentally incapable of caring for himself,

in the case of any taxable year beginning in such calendar year, such child shall be treated as a qualifying individual described in subparagraph (A) or (B) of subsection (b)(1) (whichever is appropriate) with respect to the custodial parent (as defined in section 152(e)(4)(A)), and shall not be treated as a qualifying individual with respect to the noncustodial parent.

**(6) Payments to related individuals**

- No credit shall be allowed under subsection (a) for any amount paid by the taxpayer to an individual-
- (A) with respect to whom, for the taxable year, a deduction under section 151(c) (relating to deduction for personal exemptions for dependents) is allowable either to the taxpayer or his spouse, or
  - (B) who is a child of the taxpayer (within the meaning of section 152(f)(1)) who has not attained the age of 19 at the close of the taxable year.

For purposes of this paragraph, the term "taxable year" means the taxable year of the taxpayer in which the service is performed.

**(7) Student**

The term "student" means an individual who during each of 5 calendar months during the taxable year is a full-time student at an educational organization.

**(8) Educational organization**

The term "educational organization" means an educational organization described in section 170(b)(1)(A)(ii).

**(9) Identifying information required with respect to service provider**

- No credit shall be allowed under subsection (a) for any amount paid to any person unless-
- (A) the name, address, and taxpayer identification number of such person are included on the return claiming the credit, or
  - (B) if such person is an organization described in section 501(c)(3) and exempt from tax under section 501(a), the name and address of such person are included on the return claiming the credit.

In the case of a failure to provide the information required under the preceding sentence, the preceding sentence shall not apply if it is shown that the taxpayer exercised due diligence in attempting to provide the information so required.

**(10) Identifying information required with respect to qualifying individuals**

No credit shall be allowed under this section with respect to any qualifying individual unless the TIN of such individual is included on the return claiming the credit.

**(f) Regulations**

The Secretary shall prescribe such regulations as may be necessary to carry out the purposes of this section.



## §117. Qualified scholarships

### (a) General rule

Gross income does not include any amount received as a qualified scholarship by an individual who is a candidate for a degree at an educational organization described in section 170(b)(1)(A)(ii).

### (b) Qualified scholarship

For purposes of this section-

#### (1) In general

The term "qualified scholarship" means any amount received by an individual as a scholarship or fellowship grant to the extent the individual establishes that, in accordance with the conditions of the grant, such amount was used for qualified tuition and related expenses.

#### (2) Qualified tuition and related expenses

For purposes of paragraph (1), the term "qualified tuition and related expenses" means-

(A) tuition and fees required for the enrollment or attendance of a student at an educational organization described in section 170(b)(1)(A)(ii), and

(B) fees, books, supplies, and equipment required for courses of instruction at such an educational organization.

### (c) Limitation

#### (1) In general

Except as provided in paragraph (2), subsections (a) and (d) shall not apply to that portion of any amount received which represents payment for teaching, research, or other services by the student required as a condition for receiving the qualified scholarship or qualified tuition reduction.

#### (2) Exceptions

Paragraph (1) shall not apply to any amount received by an individual under-

(A) the National Health Service Corps Scholarship Program under section 338A(g)(1)(A) of the Public Health Service Act,

(B) the Armed Forces Health Professions Scholarship and Financial Assistance program under subchapter I of chapter 105 of title 10, United States Code, or

(C) a comprehensive student work-learning-service program (as defined in section 448(e) of the Higher Education Act of 1965) operated by a work college (as defined in such section).

### (d) Qualified tuition reduction

#### (1) In general

Gross income shall not include any qualified tuition reduction.

#### (2) Qualified tuition reduction

For purposes of this subsection, the term "qualified tuition reduction" means the amount of any reduction in tuition provided to an employee of an organization described in section 170(b)(1)(A)(ii) for the education (below the graduate level) at such organization (or another organization described in section 170(b)(1)(A)(ii)) of-

(A) such employee, or

(B) any person treated as an employee (or whose use is treated as an employee use) under the rules of section 132(h).

#### (3) Reduction must not discriminate in favor of highly compensated, etc.

Paragraph (1) shall apply with respect to any qualified tuition reduction provided with respect to any highly compensated employee only if such reduction is available on substantially the same terms to each member of a group of employees which is defined under a reasonable classification set up by the employer which does not discriminate in favor of highly compensated employees (within the meaning of section 414(q)). For purposes of this paragraph, the term "highly compensated employee" has the meaning given such term by section 414(q).

**[(4) Repealed. Pub. L. 101-140, title II, §203(a)(1), (2), Nov. 8, 1989, 103 Stat. 830 ]**

#### (5) Special rules for teaching and research assistants

In the case of the education of an individual who is a graduate student at an educational organization described in section 170(b)(1)(A)(ii) and who is engaged in teaching or research activities for such organization, paragraph (2) shall be applied as if it did not contain the phrase "(below the graduate level)".

## §132. Certain fringe benefits

### (a) Exclusion from gross income

Gross income shall not include any fringe benefit which qualifies as a-

- (1) no-additional-cost service,
- (2) qualified employee discount,
- (3) working condition fringe,
- (4) de minimis fringe,
- (5) qualified transportation fringe,
- (6) qualified moving expense reimbursement,
- (7) qualified retirement planning services, or
- (8) qualified military base realignment and closure fringe.

sections (b) – (g) not shown

### (h) Certain individuals treated as employees for purposes of subsections (a)(1) and (2)

For purposes of paragraphs (1) and (2) of subsection (a)-

#### (1) Retired and disabled employees and surviving spouse of employee treated as employee

With respect to a line of business of an employer, the term "employee" includes-

- (A) any individual who was formerly employed by such employer in such line of business and who separated from service with such employer in such line of business by reason of retirement or disability, and
- (B) any widow or widower of any individual who died while employed by such employer in such line of business or while an employee within the meaning of subparagraph (A).

#### (2) Spouse and dependent children

##### (A) In general

Any use by the spouse or a dependent child of the employee shall be treated as use by the employee.

##### (B) Dependent child

For purposes of subparagraph (A), the term "dependent child" means any child (as defined in section 152(f)

(1) of the employee-

- (i) who is a dependent of the employee, or
- (ii) both of whose parents are deceased and who has not attained age 25.

For purposes of the preceding sentence, any child to whom section 152(e) applies shall be treated as the dependent of both parents.

sections (i) – (o) not shown

## §152. Dependent defined

### (a) In general

For purposes of this subtitle, the term "dependent" means-

- (1) a qualifying child, or
- (2) a qualifying relative.

### (b) Exceptions

For purposes of this section-

#### (1) Dependents ineligible

If an individual is a dependent of a taxpayer for any taxable year of such taxpayer beginning in a calendar year, such individual shall be treated as having no dependents for any taxable year of such individual beginning in such calendar year.

#### (2) Married dependents

An individual shall not be treated as a dependent of a taxpayer under subsection (a) if such individual has made a joint return with the individual's spouse under section 6013 for the taxable year beginning in the calendar year in which the taxable year of the taxpayer begins.

#### (3) Citizens or nationals of other countries

##### (A) In general

The term "dependent" does not include an individual who is not a citizen or national of the United States unless such individual is a resident of the United States or a country contiguous to the United States.

##### (B) Exception for adopted child

Subparagraph (A) shall not exclude any child of a taxpayer (within the meaning of subsection (f)(1)(B)) from the definition of "dependent" if-

- (i) for the taxable year of the taxpayer, the child has the same principal place of abode as the taxpayer and is a member of the taxpayer's household, and
- (ii) the taxpayer is a citizen or national of the United States.

### (c) Qualifying child

For purposes of this section-

#### (1) In general

The term "qualifying child" means, with respect to any taxpayer for any taxable year, an individual-

- (A) who bears a relationship to the taxpayer described in paragraph (2),
- (B) who has the same principal place of abode as the taxpayer for more than one-half of such taxable year,
- (C) who meets the age requirements of paragraph (3),
- (D) who has not provided over one-half of such individual's own support for the calendar year in which the taxable year of the taxpayer begins, and
- (E) who has not filed a joint return (other than only for a claim of refund) with the individual's spouse under section 6013 for the taxable year beginning in the calendar year in which the taxable year of the taxpayer begins.

#### (2) Relationship

For purposes of paragraph (1)(A), an individual bears a relationship to the taxpayer described in this paragraph if such individual is-

- (A) a child of the taxpayer or a descendant of such a child, or
- (B) a brother, sister, stepbrother, or stepsister of the taxpayer or a descendant of any such relative.

#### (3) Age requirements

##### (A) In general

For purposes of paragraph (1)(C), an individual meets the requirements of this paragraph if such individual is younger than the taxpayer claiming such individual as a qualifying child and-

- (i) has not attained the age of 19 as of the close of the calendar year in which the taxable year of the taxpayer begins, or
- (ii) is a student who has not attained the age of 24 as of the close of such calendar year.

##### (B) Special rule for disabled

In the case of an individual who is permanently and totally disabled (as defined in section 22(e)(3)) at any time during such calendar year, the requirements of subparagraph (A) shall be treated as met with respect to such individual.

**(4) Special rule relating to 2 or more who can claim the same qualifying child****(A) In general**

Except as provided in subparagraphs (B) and (C), if (but for this paragraph) an individual may be claimed as a qualifying child by 2 or more taxpayers for a taxable year beginning in the same calendar year, such individual shall be treated as the qualifying child of the taxpayer who is-

- (i) a parent of the individual, or
- (ii) if clause (i) does not apply, the taxpayer with the highest adjusted gross income for such taxable year.

**(B) More than 1 parent claiming qualifying child**

If the parents claiming any qualifying child do not file a joint return together, such child shall be treated as the qualifying child of-

- (i) the parent with whom the child resided for the longest period of time during the taxable year, or
- (ii) if the child resides with both parents for the same amount of time during such taxable year, the parent with the highest adjusted gross income.

**(C) No parent claiming qualifying child**

If the parents of an individual may claim such individual as a qualifying child but no parent so claims the individual, such individual may be claimed as the qualifying child of another taxpayer but only if the adjusted gross income of such taxpayer is higher than the highest adjusted gross income of any parent of the individual.

**(d) Qualifying relative**

For purposes of this section-

**(1) In general**

The term "qualifying relative" means, with respect to any taxpayer for any taxable year, an individual-

- (A) who bears a relationship to the taxpayer described in paragraph (2),
- (B) whose gross income for the calendar year in which such taxable year begins is less than the exemption amount (as defined in section 151(d)),
- (C) with respect to whom the taxpayer provides over one-half of the individual's support for the calendar year in which such taxable year begins, and
- (D) who is not a qualifying child of such taxpayer or of any other taxpayer for any taxable year beginning in the calendar year in which such taxable year begins.

**(2) Relationship**

For purposes of paragraph (1)(A), an individual bears a relationship to the taxpayer described in this paragraph if the individual is any of the following with respect to the taxpayer:

- (A) A child or a descendant of a child.
- (B) A brother, sister, stepbrother, or stepsister.
- (C) The father or mother, or an ancestor of either.
- (D) A stepfather or stepmother.
- (E) A son or daughter of a brother or sister of the taxpayer.
- (F) A brother or sister of the father or mother of the taxpayer.
- (G) A son-in-law, daughter-in-law, father-in-law, mother-in-law, brother-in-law, or sister-in-law.
- (H) An individual (other than an individual who at any time during the taxable year was the spouse, determined without regard to section 7703, of the taxpayer) who, for the taxable year of the taxpayer, has the same principal place of abode as the taxpayer and is a member of the taxpayer's household.

**(3) Special rule relating to multiple support agreements**

For purposes of paragraph (1)(C), over one-half of the support of an individual for a calendar year shall be treated as received from the taxpayer if-

- (A) no one person contributed over one-half of such support,
- (B) over one-half of such support was received from 2 or more persons each of whom, but for the fact that any such person alone did not contribute over one-half of such support, would have been entitled to claim such individual as a dependent for a taxable year beginning in such calendar year,
- (C) the taxpayer contributed over 10 percent of such support, and
- (D) each person described in subparagraph (B) (other than the taxpayer) who contributed over 10 percent of such support files a written declaration (in such manner and form as the Secretary may by regulations prescribe) that such person will not claim such individual as a dependent for any taxable year beginning in such calendar year.



**(4) Special rule relating to income of handicapped dependents****(A) In general**

For purposes of paragraph (1)(B), the gross income of an individual who is permanently and totally disabled (as defined in section 22(e)(3)) at any time during the taxable year shall not include income attributable to services performed by the individual at a sheltered workshop if-

- (i) the availability of medical care at such workshop is the principal reason for the individual's presence there, and
- (ii) the income arises solely from activities at such workshop which are incident to such medical care.

**(B) Sheltered workshop defined**

For purposes of subparagraph (A), the term "sheltered workshop" means a school-

- (i) which provides special instruction or training designed to alleviate the disability of the individual, and
- (ii) which is operated by an organization described in section 501(c)(3) and exempt from tax under section 501(a), or by a State, a possession of the United States, any political subdivision of any of the foregoing, the United States, or the District of Columbia.

**(5) Special rules for support**

For purposes of this subsection-

- (A) payments to a spouse which are includible in the gross income of such spouse under section 71 or 682 shall not be treated as a payment by the payor spouse for the support of any dependent, and
- (B) in the case of the remarriage of a parent, support of a child received from the parent's spouse shall be treated as received from the parent.

**(e) Special rule for divorced parents, etc.****(1) In general**

Notwithstanding subsection (c)(1)(B), (c)(4), or (d)(1)(C), if-

- (A) a child receives over one-half of the child's support during the calendar year from the child's parents-
  - (i) who are divorced or legally separated under a decree of divorce or separate maintenance,
  - (ii) who are separated under a written separation agreement, or
  - (iii) who live apart at all times during the last 6 months of the calendar year, and-

(B) such child is in the custody of 1 or both of the child's parents for more than one-half of the calendar year, such child shall be treated as being the qualifying child or qualifying relative of the noncustodial parent for a calendar year if the requirements described in paragraph (2) or (3) are met.

**(2) Exception where custodial parent releases claim to exemption for the year**

For purposes of paragraph (1), the requirements described in this paragraph are met with respect to any calendar year if-

- (A) the custodial parent signs a written declaration (in such manner and form as the Secretary may by regulations prescribe) that such custodial parent will not claim such child as a dependent for any taxable year beginning in such calendar year, and
- (B) the noncustodial parent attaches such written declaration to the noncustodial parent's return for the taxable year beginning during such calendar year.

**(3) Exception for certain pre-1985 instruments****(A) In general**

For purposes of paragraph (1), the requirements described in this paragraph are met with respect to any calendar year if-

- (i) a qualified pre-1985 instrument between the parents applicable to the taxable year beginning in such calendar year provides that the noncustodial parent shall be entitled to any deduction allowable under section 151 for such child, and
- (ii) the noncustodial parent provides at least \$600 for the support of such child during such calendar year.

For purposes of this subparagraph, amounts expended for the support of a child or children shall be treated as received from the noncustodial parent to the extent that such parent provided amounts for such support.

**(B) Qualified pre-1985 instrument**

For purposes of this paragraph, the term "qualified pre-1985 instrument" means any decree of divorce or separate maintenance or written agreement-

- (i) which is executed before January 1, 1985,
- (ii) which on such date contains the provision described in subparagraph (A)(i), and
- (iii) which is not modified on or after such date in a modification which expressly provides that this paragraph shall not apply to such decree or agreement.

**(4) Custodial parent and noncustodial parent**

For purposes of this subsection-

**(A) Custodial parent**

The term "custodial parent" means the parent having custody for the greater portion of the calendar year.

**(B) Noncustodial parent**

The term "noncustodial parent" means the parent who is not the custodial parent.

**(5) Exception for multiple-support agreement**

This subsection shall not apply in any case where over one-half of the support of the child is treated as having been received from a taxpayer **under the provision of subsection (d)(3).**

**(6) Special rule for support received from new spouse of parent**

For purposes of this subsection, in the case of the remarriage of a parent, support of a child received from the parent's spouse shall be treated as received from the parent.

**(f) Other definitions and rules**

For purposes of this section-

**(1) Child defined****(A) In general**

The term "child" means an individual who is-

- (i) a son, daughter, stepson, or stepdaughter of the taxpayer, or
- (ii) an eligible foster child of the taxpayer.

**(B) Adopted child**

In determining whether any of the relationships specified in subparagraph (A)(i) or paragraph (4) exists, a legally adopted individual of the taxpayer, or an individual who is lawfully placed with the taxpayer for legal adoption by the taxpayer, shall be treated as a child of such individual by blood.

**(C) Eligible foster child**

For purposes of subparagraph (A)(ii), the term "eligible foster child" means an individual who is placed with the taxpayer by an authorized placement agency or by judgment, decree, or other order of any court of competent jurisdiction.

**(2) Student defined**

The term "student" means an individual who during each of 5 calendar months during the calendar year in which the taxable year of the taxpayer begins-

- (A) is a full-time student at an educational organization described in section 170(b)(1)(A)(ii), or
- (B) is pursuing a full-time course of institutional on-farm training under the supervision of an accredited agent of an educational organization described in section 170(b)(1)(A)(ii) or a State or political subdivision of a State.

**(3) Determination of household status**

An individual shall not be treated as a member of the taxpayer's household if at any time during the taxable year of the taxpayer the relationship between such individual and the taxpayer is in violation of local law.

**(4) Brother and sister**

The terms "brother" and "sister" include a brother or sister by the half blood.

**(5) Special support test in case of students**

**For purposes of subsections (c)(1)(D) and (d)(1)(C),** in the case of an individual who is-

- (A) a child of the taxpayer, and
- (B) a student,

amounts received as scholarships for study at an educational organization **described in section 170(b)(1)(A)(ii)** shall not be taken into account.

sections (6) and (7) not shown

## §213. Medical, dental, etc., expenses

### (a) Allowance of deduction

There shall be allowed as a deduction the expenses paid during the taxable year, not compensated for by insurance or otherwise, for medical care of the taxpayer, his spouse, or a dependent (as defined in section 152, determined without regard to subsections (b)(1), (b)(2), and (d)(1)(B) thereof), to the extent that such expenses exceed 10 percent of adjusted gross income.

### (b) Limitation with respect to medicine and drugs

An amount paid during the taxable year for medicine or a drug shall be taken into account under subsection (a) only if such medicine or drug is a prescribed drug or is insulin.

### (c) Special rule for decedents

#### (1) Treatment of expenses paid after death

For purposes of subsection (a), expenses for the medical care of the taxpayer which are paid out of his estate during the 1-year period beginning with the day after the date of his death shall be treated as paid by the taxpayer at the time incurred.

#### (2) Limitation

Paragraph (1) shall not apply if the amount paid is allowable under section 2053 as a deduction in computing the taxable estate of the decedent, but this paragraph shall not apply if (within the time and in the manner and form prescribed by the Secretary) there is filed-

- (A) a statement that such amount has not been allowed as a deduction under section 2053, and
- (B) a waiver of the right to have such amount allowed at any time as a deduction under section 2053.

### (d) Definitions

For purposes of this section-

#### (1) The term "medical care" means amounts paid-

- (A) for the diagnosis, cure, mitigation, treatment, or prevention of disease, or for the purpose of affecting any structure or function of the body,
- (B) for transportation primarily for and essential to medical care referred to in subparagraph (A),
- (C) for qualified long-term care services (as defined in section 7702B(c)), or
- (D) for insurance (including amounts paid as premiums under part B of title XVIII of the Social Security Act, relating to supplementary medical insurance for the aged) covering medical care referred to in subparagraphs (A) and (B) or for any qualified long-term care insurance contract (as defined in section 7702B(b)).

In the case of a qualified long-term care insurance contract (as defined in section 7702B(b)), only eligible long-term care premiums (as defined in paragraph (10)) shall be taken into account under subparagraph (D).

(2) AMOUNTS PAID FOR CERTAIN LODGING AWAY FROM HOME TREATED AS PAID FOR MEDICAL CARE.-Amounts paid for lodging (not lavish or extravagant under the circumstances) while away from home primarily for and essential to medical care referred to in paragraph (1)(A) shall be treated as amounts paid for medical care if-

- (A) the medical care referred to in paragraph (1)(A) is provided by a physician in a licensed hospital (or in a medical care facility which is related to, or the equivalent of, a licensed hospital), and
- (B) there is no significant element of personal pleasure, recreation, or vacation in the travel away from home.

The amount taken into account under the preceding sentence shall not exceed \$50 for each night for each individual.

(3) PRESCRIBED DRUG.-The term "prescribed drug" means a drug or biological which requires a prescription of a physician for its use by an individual.

(4) PHYSICIAN.-The term "physician" has the meaning given to such term by section 1861(r) of the Social Security Act (42 U.S.C. 1395x(r)).

(5) SPECIAL RULE IN THE CASE OF CHILD OF DIVORCED PARENTS, ETC.-Any child to whom section 152(e) applies shall be treated as a dependent of both parents for purposes of this section.

Sections (6) – (10) not shown

### (e) Exclusion of amounts allowed for care of certain dependents

Any expense allowed as a credit under section 21 shall not be treated as an expense paid for medical care.

### (f) Special Rules for 2013 Through 2018

In the case of any taxable year-

- (1) beginning after December 31, 2012, and ending before January 1, 2017, in the case of a taxpayer if such taxpayer or such taxpayer's spouse has attained age 65 before the close of such taxable year, and
- (2) beginning after December 31, 2016, and ending before January 1, 2019, in the case of any taxpayer,

subsection (a) shall be applied with respect to a taxpayer by substituting "7.5 percent" for "10 percent".

**§7703. Determination of marital status****(a) General rule**

For purposes of part V of subchapter B of chapter 1 and those provisions of this title which refer to this subsection-

(1) the determination of whether an individual is married shall be made as of the close of his taxable year; except that if his spouse dies during his taxable year such determination shall be made as of the time of such death; and

(2) an individual legally separated from his spouse under a decree of divorce or of separate maintenance shall not be considered as married.

**(b) Certain married individuals living apart**

For purposes of those provisions of this title which refer to this subsection, if-

(1) an individual who is married (within the meaning of subsection (a)) and who files a separate return maintains as his home a household which constitutes for more than one-half of the taxable year the principal place of abode of a child (within the meaning of section 152(f)(1)) with respect to whom such individual is entitled to a deduction for the taxable year under section 151 (or would be so entitled but for section 152(e)),

(2) such individual furnishes over one-half of the cost of maintaining such household during the taxable year, and

(3) during the last 6 months of the taxable year, such individual's spouse is not a member of such household,

such individual shall not be considered as married.



## Appendix C

### Provisions per Question and Data Tuples

Table C1. Provisions per Question  
Blue Scenario  
Gold Scenario

Table C2. Question Data Tuples  
Blue Scenario  
Gold Scenario

Notes:

Table abbreviation CR = cross-reference

Table C1.

<i>Provisions per question Blue scenario</i>								
<b>Q</b>	<b>Section</b>		<b>Tokens</b>	<b>Nodes</b>	<b>Total CR</b>	<b>Support CR</b>	<b>Scope CR</b>	
<b>Q1</b>	§117(a)		33	1	1	1	0	
	§117(b)		110	5	1	1	0	
	§117(c)(1)		50	2	3	0	3	
	§117(c)(2); 117(d)	Heading only	6	2				
		Q1 totals		199	10	5	2	3
<b>Q2</b>	§152(a)		21	3	0	0	0	
	§152(c)(1-3(A))		270	14	5	5	0	
	§152(f)(5)		53	4	3	1	2	
		Q2 totals		344	21	8	6	2
<b>Q3</b>	§152(e)(1)		132	7	4	2	2	
	§152(c)(1)(B)		20	1	0	0	0	
	§152(d)(1)(C)	Heading only	3	1				
	§152(e)(2)		96	3	1	1	0	
	§152(e)(3)	Heading only	6	1				
	§152(e)(4)		47	3	0	0	0	
	Q3 totals		304	16	5	3	2	
<b>Q4</b>	§7703(a)		91	3	0	0	0	
	§7703(b)		141	4	4	3	1	
	§152(f)(1)(A)		43	5	0	0	0	
	§152(f)(1)(B)	Heading only	3	1				
	§152(f)(1)(C)	Heading only	4	1				
		Q4 totals		282	14	4	3	1
<b>Q5</b>	§2(b)(1)		207	8	4	1	3	
	§2(a)(1)		37	3	0	0	0	
	§152(b)(2)		51	2	2	1	1	
	§152(b)(3)(A)	Heading only	7	1				
	§152(b)(3)(B)	Heading only	5	1				
	§152(e)	Heading only	7	1				
		Q5 totals		314	16	6	2	4

						<b>Total</b>	<b>Support</b>	<b>Scope</b>
<b>Q</b>	<b>Section</b>			<b>Tokens</b>	<b>Nodes</b>	<b>CR</b>	<b>CR</b>	<b>CR</b>
<b>Q6</b>	§2(c)			38	1	1	1	0
			Q6 totals	38	1	1	1	0
<b>Q7</b>	§213(f)			86	3	1	0	1
			Q7 totals	86	3	1	0	1
<b>Q8</b>	§213(a)			64	1	4	1	3
	§152(b)(1)			52	2	0	0	0
	§152(b)(2)		Heading only	3	2			
	§152(d)(1)(B)		Heading only	3	1			
	§213(d)(5)			32	2	1	0	1
			Q8 totals	154	8	5	1	4
<b>Q9</b>	§213(a)			64	1	4	1	3
	§152(b)(1)			52	2	0	0	0
	§152(b)(2)		Heading only	3	2			
	§152(d)(1)(B)		Heading only	3	1			
	§213(d)(5)			32	2	1	0	1
			Q9 totals	154	8	5	1	4
<b>Q10</b>	§152(b)(2)			51	2	2	1	1
	§213(a)			64	1	4	1	3
			Q10 totals	115	3	6	2	4

Table C1 (continued).

<i>Provisions per question Gold Scenario</i>								
<b>Q</b>	<b>Section</b>		<b>Tokens</b>	<b>Nodes</b>	<b>Total CR</b>	<b>Support CR</b>	<b>Scope CR</b>	
<b>Q1</b>	§152(d)(1)		111	6	2	2	0	
	§152(d)(2)		41	2	2	1	1	
	§152(d)(3)(A)		44	2	2	1	1	
	§152(d)(3)(C)		11	1	0	0	0	
		Q1 totals	207	11	6	4	2	
<b>Q2</b>	§152(e)(1)		132	7	4	2	2	
	§152(c)(1)(B)		20	1	0	0	0	
	§152(d)(1)(C)	Heading only	3	1				
	§152(e)(2)		96	3	1	1	0	
	§152(e)(3)	Heading only	6	1				
	§152(e)(4)		47	3	0	0	0	
	§152(e)(5)		37	1	1	0	1	
	§152(d)(3)	Heading only	8	1				
		Q2 totals	349	18	6	3	3	
<b>Q3</b>	§2(c)		38	1	1	1	0	
	§2(b)(1)		207	8	4	1	3	
	§2(a)(1)		37	3	0	0	0	
	§152(e)	Heading only	7	1				
	§152(b)(2)		51	2	2	1	1	
	§152(b)(3)(A) & (B)		12	2				
	§2(b)(3)		69	5	3	0	3	
	§152(d)(2)(H)		53	3	1	0	1	
	§152(d)(3)	Heading only	8	1				
		Q3 totals	482	26	11	3	8	

Table C1 (continued)

<i>Provisions per question Gold Scenario</i>								
<b>Q</b>	<b>Section</b>		<b>Tokens</b>	<b>Nodes</b>	<b>CR</b>	<b>Support CR</b>	<b>Scope CR</b>	
<b>Q4</b>	§117(d)(1-2)		93	5	1	1	0	
	§132(h)		133	5	4	3	1	
	§132(a)(1)		20	2	0	0	0	
	152(f)(1)(A)		43	5	0	0	0	
		Q4 totals	289	17	5	4	1	
<b>Q5</b>	§117(c)(1)		48	2	3	0	3	
	§117(c)(2)	Heading only	2	1				
	§117(a)	Heading only	3	1				
	§117(d)(2)		51	1				
	§117(d)(5)		59	1	1	0	1	
		Q5 totals	163	6	4	0	4	
<b>Q6</b>	§21(a)		4	1	0	0	0	
	§21(a)(1)		70	2	2	2	0	
	§21(a)(2)		45	2	1	1	0	
	§21(b)		13	1	0	0	0	
	§21(b)(1)		118	2	5	2	3	
	§21(b)(2)	Heading only	3					
	§21(b)(2)(A)		84	3	0	0	0	
	§21(c)		66	3	1	0	1	
	§21(e)(5)		103	2	4	3	1	
		Q6 totals	506	16	13	8	5	
<b>Q7</b>	§152(d)(1)(B)		25	3	1	1	0	
	§21(b)(1)(B)		53	3	4	1	3	
		Q7 totals	78	6	5	2	3	
<b>Q8</b>	§21(e)(5)		103	2	4	3	1	
	§152(e)(4)(A)		20	3	0	0	0	
		Q8 totals	123	5	4	3	1	

Table C2.

<i>Question data tuples - Blue scenario</i>									
			Total	Total	Support	Scope	Total	Support	Scope
Q	Tokens	Nodes	CR	CR/Tokens	.Tokens	.Tokens	CR/Nodes	.Nodes	.Nodes
Q1	199	10	5	0.628	0.251	0.377	0.500	0.200	0.300
Q2	344	21	8	0.581	0.436	0.145	0.381	0.286	0.095
Q3	304	16	5	0.411	0.247	0.164	0.313	0.188	0.125
Q4	282	14	4	0.355	0.266	0.089	0.286	0.214	0.071
Q5	314	16	6	0.478	0.159	0.318	0.375	0.125	0.250
Q6	38	1	1	0.658	0.658	0.000	1.000	1.000	0.000
Q7	86	3	1	0.291	0.000	0.291	0.333	0.000	0.333
Q8	154	8	5	0.812	0.162	0.649	0.625	0.125	0.500
Q9	153	5	5	0.812	0.162	0.649	0.625	0.125	0.500
Q10	115	3	6	1.304	0.435	0.870	2.000	0.667	1.333

Table C2 (continued).

<i>Question data tuples - Gold Scenario</i>									
			Total	Total	Support	Scope	Total	Support	Scope
Q	Tokens	Nodes	CR	CR/Tokens	.Tokens	.Tokens	CR/Nodes	.Nodes	.Nodes
Q1	207	11	6	0.725	0.483	0.242	0.545	0.364	0.182
Q2	349	18	6	0.430	0.215	0.215	0.333	0.167	0.167
Q3	482	26	11	0.571	0.156	0.415	0.423	0.115	0.308
Q4	289	17	5	0.433	0.346	0.087	0.294	0.235	0.059
Q5	163	5	4	0.613	0.000	0.613	0.800	0.000	0.800
Q6	506	16	13	0.642	0.395	0.247	0.813	0.500	0.313
Q7	78	6	5	1.603	0.641	0.962	0.833	0.333	0.500
Q8	123	5	4	0.813	0.610	0.203	0.800	0.600	0.200

## **Appendix D**

### **IRB Approvals**

Rochester Institute of Technology

State University of New York  
(approval addressed to faculty sponsor)

Nova Southeastern University



**R·I·T****Rochester Institute of Technology**

RIT Institutional Review Board for the  
Protection of Human Subjects in Research  
141 Lomb Memorial Drive  
Rochester, New York 14623-5604  
Phone: 585-475-7673  
Fax: 585-475-7990  
Email: hmfsrs@rit.edu

**Form C**  
**IRB Decision Form**  
**FWA# 0000731**

**TO:** Jeffery Lasky  
**FROM:** RIT Institutional Review Board  
**DATE:** **October 16, 2018**  
**RE:** Decision of the RIT Institutional Review Board

Project Title – Using the Internal Revenue Code

The Institutional Review Board (IRB) has taken the following action on your project named above.

Exempt 46.101 (b) (1)

Now that your project is approved, you may proceed as you described in the Form A.

You are required to submit to the IRB any:

- **Proposed** modifications and wait for approval before implementing them,
- Unanticipated risks, and
- Actual injury to human subjects.



Heather Foti, MPH  
Associate Director  
Office of Human Subjects Research

# GENESE0

Institutional Review Board  
for the Protection of Human Participants  
Federal Wide Assurance ID #FWA00000208

State University of New York  
1 College Circle Doty 303  
Geneseo, New York 14454  
Phone | 585.245.5547

November 5, 2018

Harry Howe  
School of Business  
South Hall 107-SUNY  
1 College Circle  
Geneseo, NY 14454

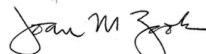
Dear Harry,

Congratulations! The Institutional Review Board (IRB) has reviewed your research project (IRB code #201819018, 'Using the Internal Revenue Code'), and it is approved, effective immediately. This approval expires in one year, on November 5, 2019.

As Principal Investigator for this study, you are required by federal regulations to inform the IRB of any proposed changes to your research that will affect human participants. Changes should be reviewed and approved before they are initiated. Unanticipated problems and adverse events should be reported to the IRB as they occur.

Thank you for this submission to the IRB and we wish the best for you and your participants. Should you have any questions regarding human participants, please feel free to contact me.

Sincerely,



Joan Zook, Chair  
Institutional Review Board for the Protection of Human Participants  
585-245-5033  
zook@geneseo.edu



MEMORANDUM

To: **Jeffrey Lasky**

From: **Ling Wang, Ph.D.,  
Center Representative, Institutional Review Board**

Date: **November 6, 2018**

Re: **IRB #: 2018-574; Title, "Using the Internal Revenue Code"**

---

I have reviewed the above-referenced research protocol at the center level. Based on the information provided, I have determined that this study is exempt from further IRB review under **45 CFR 46.101(b) (Exempt 1: Educational research in educational settings)**. You may proceed with your study as described to the IRB. As principal investigator, you must adhere to the following requirements:

- 1) **CONSENT:** If recruitment procedures include consent forms, they must be obtained in such a manner that they are clearly understood by the subjects and the process affords subjects the opportunity to ask questions, obtain detailed answers from those directly involved in the research, and have sufficient time to consider their participation after they have been provided this information. The subjects must be given a copy of the signed consent document, and a copy must be placed in a secure file separate from de-identified participant information. Record of informed consent must be retained for a minimum of three years from the conclusion of the study.
- 2) **ADVERSE EVENTS/UNANTICIPATED PROBLEMS:** The principal investigator is required to notify the IRB chair and me (954-262-5369 and Ling Wang, Ph.D., respectively) of any adverse reactions or unanticipated events that may develop as a result of this study. Reactions or events may include, but are not limited to, injury, depression as a result of participation in the study, life-threatening situation, death, or loss of confidentiality/anonymity of subject. Approval may be withdrawn if the problem is serious.
- 3) **AMENDMENTS:** Any changes in the study (e.g., procedures, number or types of subjects, consent forms, investigators, etc.) must be approved by the IRB prior to implementation. Please be advised that changes in a study may require further review depending on the nature of the change. Please contact me with any questions regarding amendments or changes to your study.

The NSU IRB is in compliance with the requirements for the protection of human subjects prescribed in Part 46 of Title 45 of the Code of Federal Regulations (45 CFR 46) revised June 18, 1991.

Cc: Amon Seagull  
Ling Wang, Ph.D.

## Appendix E

### Scenario Solutions

Question	Solution and explanation
<p><b><u>Blue Scenario</u></b></p>	
<p><b>Q1.</b> For 2018, what is the amount of Child's gross income? §117(a), (b), (c)(1)</p>	<p><b>\$3,000</b> Qualified scholarship amounts are excluded from gross income (§117(a)) to the extent that the scholarship funds are used for qualified tuition and related expenses (§117(b)(1)).</p> <p>Qualified tuition and related expenses are defined in §117(b)(2). Child's qualified expenses were: \$14,000 tuition; \$1,000 fees; \$1,000 textbooks; \$1,000 lab supplies. Total = \$17,000. Meal expenses are not qualified expenses. Of the \$18,000 scholarship award, Child applied \$17,000 to qualified tuition and related expenses. The difference of <b>\$1,000</b> is includable in gross income.</p> <p>The <b>\$2,000</b> stipend is not excluded from gross income (§117(c)(1)).</p> <p>Child's gross income = \$1,000 + \$2,000 = <b>\$3,000</b></p>
<p><b>Q2.</b> Is Child a dependent of Parent 1 by virtue of being a qualifying child of Parent 1? §152(a) §152(c), but not [§152(c)(3)(B), c(4)] §152(f)(5)</p>	<p><b>Yes.</b> §152(a) defines a dependent as a qualifying child or a qualifying relative.</p> <p>§152(c)(1), c(2), and c(3) are the eligibility criteria for being considered a qualifying child: <b>relationship</b> to Parent 1 is child [c(2)(A)]; <b>residency</b>: principal place of abode is same as Parent 1 [c(1)(B)]; meets the <b>age requirement</b> of being a student and age less than 24 [(c)(1)(C)], c(3)(A)(ii); has not provided over one-half of <b>own support</b>*</p>

	<p>[c(1)(D)];and <b>has not filed a joint return</b> [c(1)(E)]. Child meets all eligibility requirements to be considered a qualifying child of Parent 1.</p> <p>*Calculation of support 152(f)(5) excludes amounts received as scholarships when determining student contribution to own support.</p> <p>Total support 40,000 One-half. 20,000 Contribution from savings. 7,000 Stipend. <span style="float:right">2,000</span> Scholarship amount not excluded 1,000 7,000 + 2,000 +1,000 is not &gt; 20,000</p>
<p><b>Q3.</b> Do the Parents meet the provisions of §152(e)(1) so that Parent 1 can release to Parent 2 the right to claim Child as a dependent for tax year 2018? §152(e)(4)</p>	<p><b>Yes.</b> §152(e)(1) includes exceptions to the qualifying criteria for dependent status expressed as notwithstanding (c)(1)(B), (c)(4), or (d)(1)(C). (c)(4) is special tie-breaking rules for who can claim a qualifying child as a dependent when a child can be claimed by two or more individuals (N/A); d(1)(C) applies only to a qualifying relative (N/A); (c)(1)(B) is the requirement that the Child have the same principal place of abode of taxpayer.</p> <p>The remaining requirements are also met. Child receives over one-half support from Parents e(1)(A); Parents have lived apart for the last six months of the [2018] calendar year (e(1)(A)(iii); Child is in custody of Parent 1 for more than one-half of the calendar year.</p>
<p><b>Q4.</b> What is the marital status of Parent 1? §7703</p>	<p><b>Not considered as married.</b> The Tax Code distinguishes between being legally married and not being considered married [for tax provision purposes]. §7703 is Determination of marital status.</p>

	<p>For an individual to be considered as not married, the individual must be either legally separated (decree of divorce or of separate maintenance (a)(2)), or meets the requirements of (b)(1), b(2), and b(3). (b)(1), Parent 1 is married; Parent 1 files a separate return; Parent 1 maintains a household which is the principal place of abode of a child for whom Parent 1 is entitled to a deduction or would be entitled but for §152(e)(1) (see Q3). b(2), Parent 1 provides more than one-half of the cost of maintaining the household; and b(3), Parent 2 is not a member of the household during the last six months of the year.</p> <p>Parent 1 meets all requirements for being considered not married.</p>
<p><b>Q5.</b> Do the §2(b)(1) provisions support Parent 1's eligibility to be considered a head of household?</p>	<p><b>Yes.</b>  2(b)(1): Parent 1 is not married (determined in Q4 above), and is not a surviving spouse; 2(b)(1)(A): Parent 1 maintains a home and household which is the principle place of abode of 2(b)(1)(A)(i): a qualifying child (Child) determined without regard to 152(e) where Parent 1 released to Parent 2 the right to claim Child as a dependent.</p> <p>2(b)(1)(A)(i)(I) and (II) are conditions that disqualify a child from being considered a qualifying child: (I) the child is married, and (II), the child files a joint return with spouse, (§152(b)(2), or child is not a U.S. citizen (N/A) (§152(b)(3). Child files a separate 2018 return.</p> <p>Parent 1 provides more than half the cost of maintaining the household.</p>
<p><b>Q6.</b> Does the §2(c) provision support Parent 1's eligibility to being considered a head of household?</p>	<p><b>Yes.</b>  §2(c) contains a cross-reference to §7703(b), Certain married individuals living apart. §7703 was considered in Q4, where it is determined that Parent 1's</p>

	marital status is: not considered as married.
<p><b>Q7.</b> In certain cases, out-of-pocket medical care expenses are deductible to the extent such expenses exceed a certain percent of adjusted gross income. For the 2018 tax year, what is the applicable percent amount? §213(f)</p>	<p><b>7.5%</b> §213(f) is Special rules for 2013 through 2018.</p> <p>§213(f)(1) expired 12/31/16 (N/A). For tax years 2018 and 2018, §213(f)(2) adjusts the §213(a) percentage from 10% to 7.5% of AGI.</p>
<p><b>Q8.</b> What amount, if any, can Parent 1 deduct as a medical expense? §213(a) §213(d)(5)</p>	<p><b>\$1,500</b> §213(a) allows a deduction by the taxpayer (Parent 1) for unreimbursed (out-of-pocket) medical care expenses for the taxpayer, taxpayer's spouse, or <b>a dependent</b>. Parent 1 released the claim for Child as dependent to Parent 2.</p> <p>However, according to §213(d)(5), Special Rule in the Case of Divorced Parents, Etc., any child to whom §152(e) applies (see Q3 above) is treated as a dependent of both parents for the purpose of §213. Parent 1 then is eligible to deduct Child 1's medical care expenses subject to the §213(a) percentage limitation.</p> <p>AGI of Parent 1   \$60,000 7.5% threshold    4,500 Parent 1's share of medical care expenses \$6,000 Deductible amount = \$6,000 – 4,500 = \$1,500</p>
<p><b>Q9.</b> What amount, if any, can Parent 2 deduct as a medical expense? §213(a) §213(d)(5)</p>	<p><b>\$2,250</b> Parent 2 will treat Child as a dependent for 2018 by virtue of §152(e). See Q3 above.</p> <p>AGI of Parent 1   \$50,000 7.5% threshold    3,750 Parent 2's share of medical care expenses \$6,000 Deductible amount = \$6,000 – 3,750 = \$2,250</p>
<p><b>For the following question only, Child and Spouse will file a joint return for year 2018.</b></p> <p><b>Q10.</b> What amount, if any, can Parent 1 deduct as a medical expense? §152(b)(2) §213(a)</p>	<p><b>\$ 1,500 (same amount as Q8 above)</b> Medical expense deductions <b>are allowed</b> for the taxpayer, taxpayer's spouse, or <b>a dependent</b>. (§213(a)).</p> <p>Child <b>cannot be considered a dependent</b> of Parent 1 or of Parent 2 because Child has filed a joint return with Spouse (§152(b)(2)).</p>

	<p><b>However</b>, §213(a) reads in part ... or a dependent (<b>as defined in section 152 determined without regard to subsections ... (b)(2)</b>), and .... Hence, for purpose of allowing a medical expense deduction, Child's status as a dependent of Parent 1 is not reversed by §152(b)(2).</p>
--	---



<b>Gold Scenario</b>	
<p><b>Q1.</b> Is Child a dependent of Parent 1 by virtue of being a qualifying relative of Parent 1?</p> <p>§152(d)(1), (d)(2), (d)(3)(A), and (d)(3)(C), [but not §152 (d)(3)(B), (d)(3)(D), (d)(4), (d)(5)].</p> <p>Note: assume the requirements in (d)(3)(B) and (d)(3)(D) are satisfied.</p>	<p><b>Yes.</b> §152(d)(1), and (d)(2), are the eligibility criteria for being considered a qualifying relative: child is a qualifying relationship to taxpayer, and Child is a child d(1)(A) and d(2)(A); qualifying relative’s gross income must be less than the exemption amount defined in §151(d). For tax year 2018, the §151(d) amount is \$4,150.</p> <p>Parent 1 must provide more than one-half of Child’s support for 2018. The scenario change of circumstances describes a multiple-support agreement for Child where Parent 1 provides only 15% of Child’s support.</p> <p><b>However</b>, if the multiple support agreement meets the §152(d)(3) criteria, then over one-half support of Child will be considered to have been provided by Parent 1. §152(d)(3) “For purposes of paragraph (1)(C), over one-half support ... will be treated as received from the taxpayer if-”. All criteria are satisfied.</p> <p>The (d)(3)(A) criteria, no one person contributed over one-half of support, is met.</p> <p>The (d)(3)(B) criteria, over one-half of support was received from two or more persons, each of whom would be able to claim Child as a dependent but for the support test, is met. The two grandparents, and Parent 2 together contribute more than 50% and each could claim Child as a dependent since each has a family relationship to Child specified in (d)(2)(C).</p> <p>The (d)(3)(C) criteria is met, as Parent has Parent 1 has contributed over 10%.</p> <p>The (d)(3)(D) IRS filing requirement is met by stipulation.</p>
<p><b>Q2.</b> Do the Parents meet the provisions of §152(e)(1) so that Parent 1 can release to</p>	<p><b>No.</b> §152(e)(5) disallows the use of §152(e)(1) by an exception reference to §152(d)(3), multiple support agreements.</p>

<p>Parent 2 the right to claim Child as a dependent for tax year 2018?</p> <p>§152(e)(4), 152(e)(5)</p>	<p>According to §152(e)(5), subsection (e) will not apply if over one-half support of a [otherwise qualifying] child]is treated as having been received by taxpayer [Parent 1] under §152(d)(3).</p>
<p><b>Q3.</b> Can Parent 1 claim Head of Household filing status?</p> <p>§2(b)(1) §2(b)(3)</p>	<p><b>No.</b></p> <p>Parent 1 meets the requirements to be considered a head of household under §2(b)(1)(A)(ii), since Child is a dependent by virtue of being a qualifying relative (married but filing separately).</p> <p><b>However,</b> Parent 1 is disqualified from being considered a head of household due to the cross-reference from §2(b)(3)(B)(ii) to §152(d)(3), multiple support agreement. The disqualification is because Child would not have been a dependent but for the multiple support agreement.</p>
<p><b>Q4.</b> Under the University’s tuition reduction program, in 2018, Child’s tuition was reduced by \$8,000.</p> <p>§117(d)(1), (d)(2)</p> <p>What amount, if any of the tuition reduction is taxable income to Parent 1?</p>	<p><b>\$8,000 (the entire amount)</b></p> <p>In general, the amount of qualified tuition reduction is not included in gross income (§117(d)(1)) if the tuition reduction is for <b>undergraduate education</b> (§117(d)(2)), of the employee (d)(2)(A), or the employee’s spouse or a dependent of the employee (§2(d)(2)(B) cross reference to §132(h).</p> <p>§132 is Certain Fringe Benefits. §132(a)(1) qualifies the value of a no-additional-cost service provided by the employer to be excluded from employee’s gross income. (The scenario explains that tuition reduction is a no-additional-cost service). §132(h) includes a cross-reference to §132(a)(1)</p> <p>However, Child is a graduate student, so no amount of the tuition reduction is excluded from Parent 1’s (the employee) gross income.</p>
<p><b>Q5.</b> Under the University’s tuition reduction program, in 2018, Child’s tuition was reduced by \$8,000. In addition, the University provided a \$3,000 stipend (a payment) for Child’s required service as a teaching assistant.</p> <p>§117(d)(1), (2) §117(d)(5)</p>	<p><b>\$0.</b></p> <p>§117(d)(5) is Special Rules for Teaching and Research Assistants. Since Child is providing teaching assistant services as a graduate student, §117(d)(5) is applicable. This provision modifies the text of §117(d)(2) to delete the phrase “below the graduate level”.</p> <p>In this case, §117(d)(2) applies identically to undergraduate and graduate level education,</p>

<p>What amount, if any, of Child's award is taxable income to Parent 1?</p>	<p>and so the general rule that excludes qualified tuition reduction from gross income of Parent 1 applies to Child's use of the tuition reduction at the graduate level.</p>
<p><b>NOTE: The following is inserted immediately before Q6 in the Gold Scenario:</b></p> <p><i>Whether you answered Yes or No to Question 2, please continue assuming that the Parents did meet the requirements of the specified provisions, and that Parent 1 did release to Parent 2 the right to claim Child as a dependent.</i></p> <p><b>Q6.</b> As noted in the scenario, Parent 1 paid \$1,600 for the cost of the household and child care services that were incurred as a result of Child's illness. What amount, if any, can Parent 1 claim as a tax credit?</p> <p>§21(a)(1), (a)(2), (b)(1), b(2)(A), (c) [but not the sentence following (c)(2)]</p> <p>§21(e)(5)</p> <p>Note: assume the (a)(2) applicable percentage is 20%</p>	<p><b>\$320.</b></p> <p>§21(a)(1) allows a tax credit to a taxpayer who has paid for household and dependent care expenses so that taxpayer (Parent 1) could continue to go to work. Child required such household and dependent care services, and is a qualifying individual for the purpose of the tax credit (§21(b)(1)(B)).</p> <p>§21(a)(2) is the calculation of <i>applicable percentage</i> (AP) that is used in the determining the amount of the credit. Parent 1 AGI = \$65,000; AP = 35% reduced by 1% for each \$2,000 by which Taxpayer's AGI exceeds \$15,000; <math>(65,000-15,000)/2 = 25\%</math> reduction; tentative AP = <math>35\% - 25\% = 10\%</math>. However, the §21(a)(2) AP minimum is 20%.</p> <p>Parent 1 paid \$1,600 for household and dependent care expenses; <math>\\$1,600 \times 20\% = \\$320</math> credit.</p> <p>§21(c)(1) limits the credit to \$3,000 in the case of one qualifying individual.</p> <p>§21(e)(5) specifies that in a case where §152(e) applies, a child is qualifying individual with respect to the custodial parent.</p>
<p><b>Q7. For this question only</b>, in 2018, Child earned \$8,000 as a summer intern. As noted in the scenario, Parent 1 paid \$1,600 for the cost of the household and child care services that were incurred as a result of Child's illness. What amount, if any, can Parent 1 claim as a tax credit?</p> <p>§152(d)(1)(B) §21(b)(1)(B)</p>	<p><b>\$320.</b></p> <p>In general, an individual will not qualify to be considered a qualifying relative if the individual's gross income exceeds the exemption amount defined in §151(d) (as per §152(d)(1)(B)). For tax year 2018, the §151(d) amount is \$4,150.</p> <p>Child earned \$8,000 which is greater than \$4,150. However, §21(b)(1)(B) instructs that §152(d)(1)(B) is not to be taken into account in determining whether an individual is a dependent for purposes of §21. Child meets the other requirements to be considered a dependent, and hence is a qualifying individual for §21.</p>

<p><b>Q8.</b> As noted in the scenario, Parent 2 paid \$1,600 for the cost of the household and child care services that were incurred as a result of Child's illness. What amount, if any, can Parent 2 claim as a tax credit?</p> <p>§21(e)(5)</p>	<p><b>Zero (\$0).</b></p> <p>§21(e)(5) specifies that in a case where §152(e) applies, a child is qualifying individual with respect to the custodial parent, and <b>NOT</b> treated as a qualifying individual with respect to the noncustodial parent.</p>
--	--

## **Appendix F**

### **Tax Expert Reviewers**

**Philip C. Gelsomino II** is a Lecturer in the Department of Accounting, Saunders College of Business, Rochester Institute of Technology. He is a New York State licensed CPA with over 30 years of experience providing income tax, accounting, and auditing services to clients. Mr. Gelsomino teaches a variety of undergraduate and graduate courses in accounting, taxation, and entrepreneurship. He received both his B.S. in Accounting (1986) and M.S. in Entrepreneurship and Innovative Ventures (2018) from RIT.

**Joseph P. Grymin** is a New York State licensed CPA with 42 years of experience in individual and business taxation, estate tax planning, and related financial services. He received his BBA degree (major in accounting) from St. John Fisher College, Rochester, NY in 1970, and has remained in the Rochester area since that time. Mr. Grymin is presently owner of his own practice firm. (The researcher is a client of Mr. Grymin's firm).

## Appendix G

### Distributions of Independent Variable Values

Tokens

Nodes

Tot\_CR

Tot\_CR.Tokens

Tot\_CR.Nodes

Scope

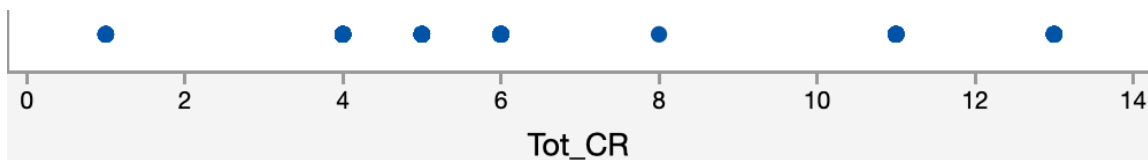
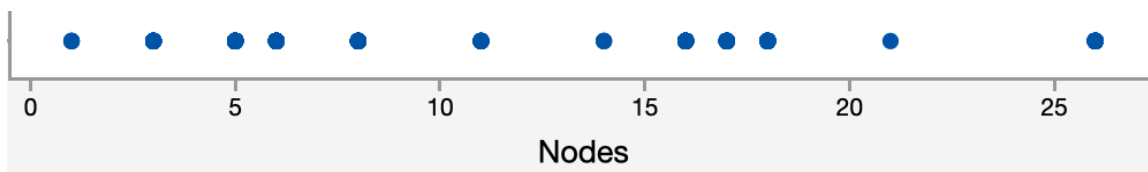
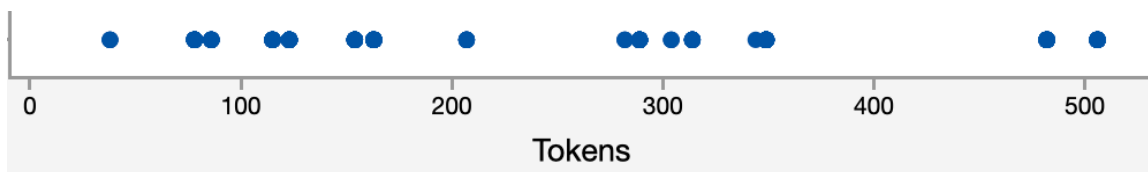
Scope.Tokens

Scope.Nodes

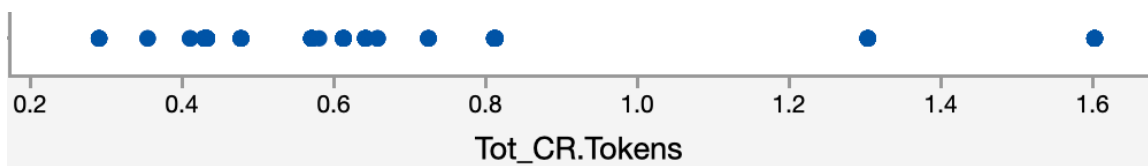
Support

Support.Tokens

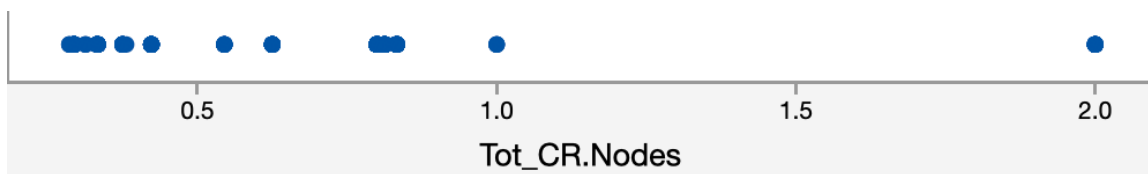
Support.Nodes



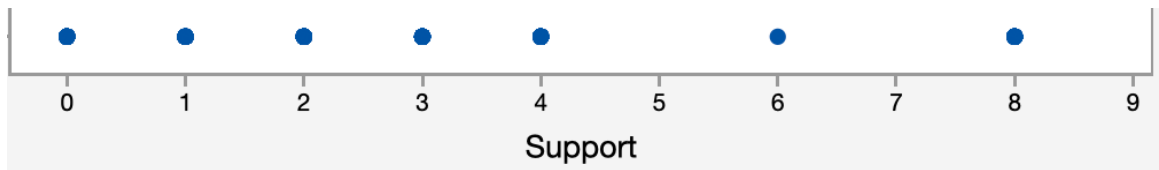
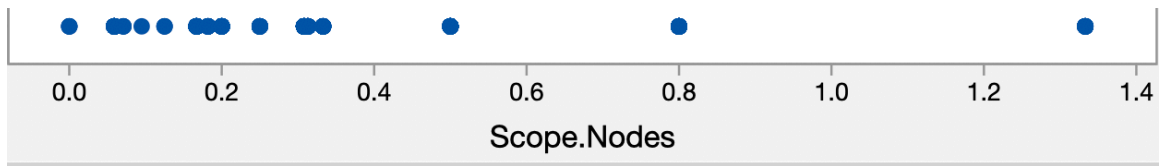
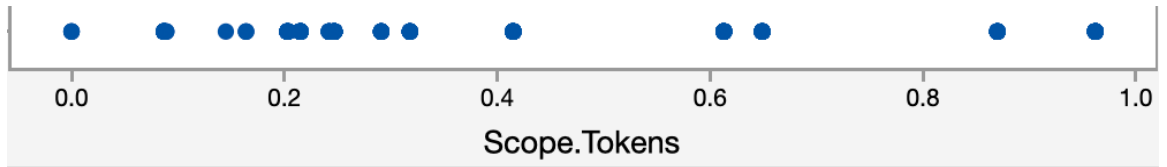
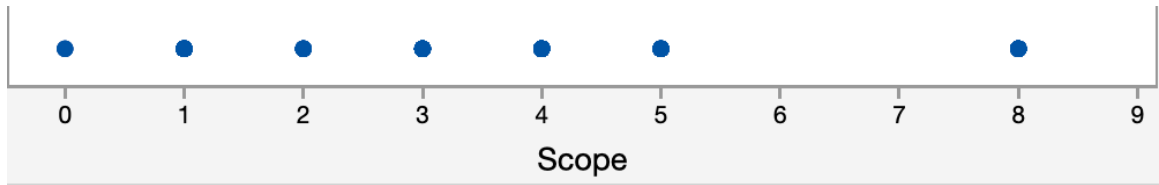
Unusual values reported



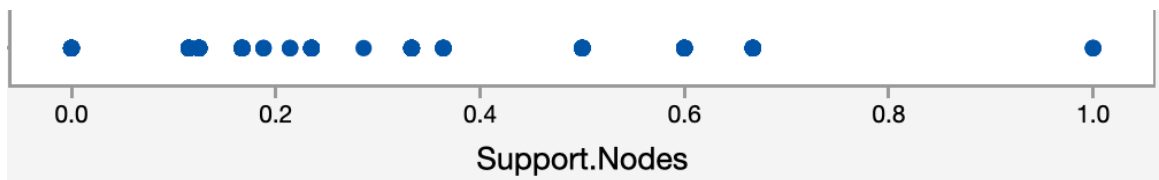
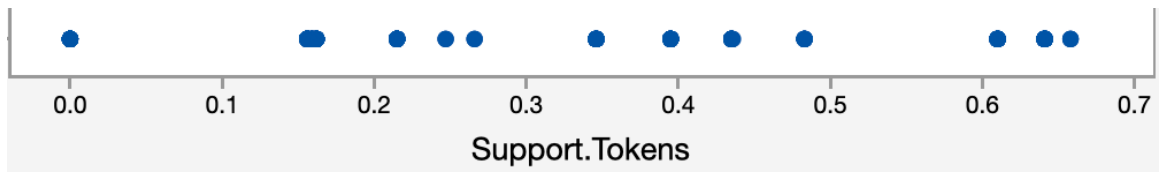
Unusual values reported



Unusual values reported



Unusual values reported



Unusual values reported



## **Appendix H**

### **Participant Study Experience Survey**

### Participant Study Experience Survey

Please circle the description that best expresses your opinion about the collection of Tax Code sections you have just read.

(1). I thought the sections were easy to understand.

Agree Very Strongly	Agree Strongly	Agree	Neither Agree nor Disagree	Disagree	Disagree Strongly	Disagree Very Strongly

(2). I thought the sections were too wordy

Agree Very Strongly	Agree Strongly	Agree	Neither Agree nor Disagree	Disagree	Disagree Strongly	Disagree Very Strongly

(3). I thought the sections were difficult to understand.

Agree Very Strongly	Agree Strongly	Agree	Neither Agree nor Disagree	Disagree	Disagree Strongly	Disagree Very Strongly

(4). I thought the language used in the sections was too technical.

Agree Very Strongly	Agree Strongly	Agree	Neither Agree nor Disagree	Disagree	Disagree Strongly	Disagree Very Strongly

(please continue to next page)

(5). I thought the sections were clearly written.

Agree Very Strongly	Agree Strongly	Agree	Neither Agree nor Disagree	Disagree	Disagree Strongly	Disagree Very Strongly

(6). I think it will be easy to remember the content of the sections tomorrow.

Agree Very Strongly	Agree Strongly	Agree	Neither Agree nor Disagree	Disagree	Disagree Strongly	Disagree Very Strongly

(7). I am confident that all or most of my answers to the scenario questions are correct.

Agree Very Strongly	Agree Strongly	Agree	Neither Agree nor Disagree	Disagree	Disagree Strongly	Disagree Very Strongly

## References

- Adedjouma, M., Sabetzadeh, M., & Briand, L. C. (2014). Automated detection and resolution of legal cross references: Approach and a study of Luxembourg's legislation. *2014 IEEE 22nd International Requirements Engineering Conference (RE)*, 63-72. <https://doi.org/10.1109/RE.2014.6912248>
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, 2, 89-195. [https://doi.org/10.1016/s0079-7421\(08\)60422-3](https://doi.org/10.1016/s0079-7421(08)60422-3)
- Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, 8, 47-89. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Barrick, J. A. (2001). The effect of code section knowledge on tax-research performance. *The Journal of the American Taxation Association*, 23(2), 20-34. <https://doi.org/10.2308/jata.2001.23.2.20>
- Benson, C. W. (2019). Section 199A: Magic dance through the labyrinth. *Washburn Law Journal*, 58, 187-216.
- Benson, R. W. (1985). The end of legalese: The game is over. *N.Y.U. Review of Law and Social Change*, 13, 519-573.
- Bhatia, V. K. (1998). Intertextuality in legal discourse. *The Language Teacher*, 22(11).
- Bhatia, V. K. (1987). Textual-mapping in British legislative writing. *World Englishes*, 6, 1-10. <https://doi.org/10.1111/j.1467-971x.1987.tb00172.x>
- Bhatia, V. K. (1983, May). *Language Studies Unit Research Report: An applied discourse analysis of English legislative writing*. Birmingham, UK: The University of Ashton in Birmingham.
- Biber, D. (2006). *University language: A corpus-based study*. Amsterdam, Netherlands: John Benjamins.
- Bowers, F. (1989). *Linguistic aspects of legislative expression*. Vancouver, Canada: University of British Columbia Press.
- Brady, D. (2019, April). *Tax reform bill made modest progress toward simplification, but significant hurdles remain*. Retrieved from National Taxpayers Union Foundation website: <https://www.ntu.org/library/doclib/2019/04/Tax-Reform-Bill-Made-Progress-Toward-Simplification-4-15.pdf>

- Breaux, T. D., & Anton, A. I. (2008). Analyzing regulatory rules for privacy and security requirements. *IEEE Transactions on Software Engineering*, 34(1), 5-20. <https://doi.org/10.1109/TSE.2007.70746>
- Breeze, R. (2013). Lexical bundles across four legal genres. *International Journal of Corpus Linguistics*, 18, 229-253. <https://doi.org/10.1075/ijcl.18.2.03bre>
- Burton, H. A., & Karlinsky, S. (2016). Complexity of tax simplification: USA perspective. In S. James, A. Sawyer, & T. Budak (Eds.), *The complexity of tax simplification: Experiences from around the world* (pp. 247-265). New York, NY: Palgrave MacMillan. [https://doi.org/10.1057/9781137478696\\_1](https://doi.org/10.1057/9781137478696_1)
- Cauble, E. (2019). Superficial proxies for simplicity in tax law. *University of Richmond Law Review*, 53, 329-372. <https://doi.org/10.2139/ssrn.3162604>
- Charrow, V. R., Crandall, J. A., & Charrow, R. P. (1982). Characteristics and functions of legal language. In R. Kittredge & J. Lehrberger (Eds.), *Sublanguage: Studies of language in restricted semantic domains* (pp. 175-190). Berlin, Germany: Walter de Gruyter.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum. <https://doi.org/10.4324/9780203774441>
- Complexity Compounded – The Internal Revenue Code [Editorial]. (1959). *Tax Executive*, 11, 99-100.
- Coven, G. E. (2000). Bad drafting: A case study of the design and implementation of the income tax subsidies for education. *Tax Lawyer*, 54, 1-64.
- Crossley, S. A., Skalicky, S., & Dascalu, M. (2019). Moving beyond classic readability formulas: New methods and new models. *Journal of Research in Reading*, 1-21. <https://doi.org/10.1111/1467-9817.12283>
- Crystal, D., & Davy, D. (1969). *Investigating English style*. Bloomington, In: Longman.
- Curtotti, M., Weibel, W., McCreath, E., Ceynowa, N., Frug, S., & Bruce, T. (2015a). Citizen science for citizen access to law. *Journal of Open Access to Law*, 3(1), 57-120.
- Curtotti, M., McCreath, E., Bruce, T., Frug, S., Weibel, W., & Ceynowa, N. (2015b). - . *Proceedings of the 15th International Conference on Artificial Intelligence and Law (ICAIL '15)*, 53-62. <https://doi.org/10.1145/2746090.2746095>

- Cyrul, W. (2013). Consistency and coherence in the "hypertext" of law: A textological approach. In M. Araszkiwicz & J. Šavelka (Eds.), *Coherence: Insights from philosophy, jurisprudence, and artificial intelligence* (Vol. 107, pp. 167-186). Dordrecht, Germany: Springer.
- Danet, B. (1985). Legal discourse. In T. A. Van Dijk (Ed.), *Handbook of discourse analysis: Discourse analysis in society* (Vol. 1, pp. 278-287). London, UK: Academic Press.
- Delaney, P. F., & Ericsson, K. A. (2016). Long-term working memory and transient storage in reading comprehension: What is the evidence? Comment on Foroughi, Werner, Barragán, and Boehm-Davis (2015). *Journal of Experimental Psychology: General*, 145, 1406-1409. <https://doi.org/10.1037/xge0000181>
- de Matt, E., Winkels, R., & Engers, T. M. (2016). Automated detection of reference structures in the law. In T. M. Van Engers (Ed.), *JURIX 2006: The Nineteenth Annual Conference: Legal knowledge and information systems* (pp. 41-50). Amsterdam, Netherlands: IOS Press.
- Dubay, W. H. (2004). The principles of readability. Retrieved from <https://files.eric.ed.gov/fulltext/ED490073.pdf>
- Ericsson, K. A., & Kintsch, W. (1995). Long term working memory. *Psychological Review*, 102, 211-245. <https://doi.org/10.1037/0033-295X.102.2.211>
- Flesch, R. (1979a). Lets Start with the Formula. In *How to write plain English: A book for lawyers and consumers* (pp. 20-32). New York, NY: Harper & Row.
- Flesch, R. (1979b). The cross-reference habit. In *How to write plain English: A book for lawyers and consumers* (pp. 81-93). New York, NY: Harper & Row.
- Foroughi, C. K., Werner, N. E., Barragan, D., & Bohm-Davis, D. A. (2015). Interruptions disrupt reading comprehension. *Journal of Experimental Psychology: General*, 144, 704-709. <https://doi.org/10.1037/xge0000074>
- Gozdz-Roszkowski, S. (2011). Multi-word patterns in legal genres. In B. Lewandowska-Tomaszczyk (Ed.), *Lodz Studies in Language: Vol. 22. Patterns of linguistic variation in American legal English: A corpus-based study* (pp. 109-143). Frankfurt am Main, Germany: Peter Lang.
- Gershon, I. R., & Maine, J. A. (2019). *A student's guide to the Internal Revenue Code* (7th ed.). Durham, NC: Carolina Academic Press.
- Gustafsson, M. (1975). *Publications of the Department of English: Some syntactic properties of English law language* (Research Report No. 4). Turku, Finland: University of Turku.

- Hamdaqa, M., & Hamou-Lhadj, A. (2011). An approach based on citation analysis to support effective handling of regulatory compliance. *Future Generation Computer Systems*, 27, 395-410. <https://doi.org/doi:10.1016/j.future.2010.09.007>
- Hand, L. (1947). Thomas Walter Swan. *Yale Law Review*, 57, 167-172
- Hart, H. L. (1994). Law as the union of primary and secondary rules. In *Clarendon Law Series: The concept of law* (2nd ed., pp. 79-99). Oxford, UK: Oxford University Press.
- Hiltunen, R. (2012). The grammar and structure of legal texts. In P. M. Tiersma & L. M. Solan (Eds.), *Oxford Handbooks in Linguistics: The Oxford handbook of language and law* (pp. 39-51). Oxford, UK: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199572120.001.0001>
- Holmes, G. E. (1926, April). Is it not time to simplify the income tax? The bulletin of the National Tax Association, 11(7), 194-195.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27, 4-21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Karlinsky, S. S., & Koch, B. S. (1987). Impact of tax law complexity on professionals. *The Journal of the American Tax Association*, 9(1), 24-34.
- Katz, D. M., & Bommarito, M. J. (2014). Measuring the complexity of the law: The United States Code. *Artificial Intelligence and Law*, 22, 337-374. <https://doi.org/10.1007/s10506-014-9160-8>
- Kittredge, R. (1982). Sublanguages (Panel). *American Journal of Computational Linguistics*, 8(2), 79-84.
- Kittredge, R., & Lehrberger, J. (1982). Introduction. In R. Kittredge & J. Lehrberger (Eds.), *Sublanguage: Studies of language in restricted semantic domains* (pp. 1-7). Berlin, Germany: Walter de Gruyter.
- Knott, A., & Dale, R. (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18, 35-62. <https://doi.org/10.1080/01638539409544883>
- Knott, A., & Mellish, C. (1996). A feature-based account of the relations signaled by sentence and clause connectives. *Language and Speech*, 39, 143-183.
- Koch, B. S., & Karlinsky, S. S. (1984). The effect of Federal income tax law reading complexity on students' task performance. *Issues in Accounting Education*, 98-110.

- Li, W. P. (2016). *Language technologies for understanding law, politics, and public policy* (Doctoral dissertation). Retrieved from <http://hdl.handle.net/1721.1/103673>
- Li, W. P., Azar, P., Larochele, D., Hill, P., & Lo, A. W. (2015). Law is code: A software engineering approach to analyzing the United States Code. *Journal of Business & Technology Law*, 10, 297-374. Retrieved from <http://digitalcommons.law.umaryland.edu/jbtl/vol10/iss2/6>
- Lötscher, A. (2008). Conceptual and textual structure in legislative texts. In A. Wagner & S. Cacciaguidi-Fahy (Eds.), *Obscurity and clarity in the law: Prospects and challenges* (pp. 129-150). Aldershot, Hampshire, England: Ashgate.
- Luttman, S. M., & Monarchi, D. E. (2015). Network analysis and visualization of the complexity of an individual tax return. *The ATA Journal of Legal Tax Research*, 13(1), 17-35. <https://doi.org/10.2308/jltr-51085>
- Maley, Y. (1987). The language of legislation. *Language in Society*, 16, 25-48. <https://doi.org/10.1017/S0047404500012112>
- Maxwell, J. C., Anton, A. I., & Earp, J. B. (2013). An empirical investigation of software engineers' ability to classify legal cross-references. *21st IEEE International Requirements Engineering Conference (RE'13)*, 24-31. <https://doi.org/10.1109/RE.2013.6636702>
- Maxwell, J. C., Anton, A. I., Swire, P., Riaz, M., & McCraw, C. M. (2012). A legal cross-references taxonomy for reasoning about compliance requirements. *Requirements Engineering*, 17(2), 99-115. <https://doi.org/10.1007/s00766-012-0152-5>
- McCaffery, E. J. (1990). The holy grail of tax simplification. *Wisconsin Law Review*, 1267-1332.
- Miller, J. A. (1993). Indeterminacy, complexity and fairness: Justifying rule simplification in the law of taxation. *Washington Law Review*, 68, 1-78.
- Miller, J. R., & Kintsch, W. (1980). Readability and recall of short prose passages: A theoretical analysis. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 335-354. <https://doi.org/10.1037/0278-7393.6.4.335>
- Morrison, A. (2014). Case law, systematic law, and a very modest suggestion. *Statute Law Review*, 35, 159-180. <https://doi.org/10.1093/slr/hmt016>
- Oei, S., & Osofsky, L. Z. (2019). Constituencies and control in statutory drafting: Interviews with government tax counsels. *Iowa Law Review*, 104, 1291-1360.



- Office of the Law Revision Counsel. (n.d.). Detailed guide to the United States Code content and features. [https://uscode.house.gov/detailed\\_guide.xhtml](https://uscode.house.gov/detailed_guide.xhtml)
- Potter, P. B., Jr. (2004). Surveying the Serbian Bog: A brief history of a judicial metaphor. *Tulane Maritime Law Journal*, 28, 519-553.
- Redish, J.C., & Selzer, J. (1985). The place of readability formulas in technical communication. *Technical Communication*, 32(4), 46-52. Retrieved from <https://www.jstor.org/stable/43094562>.
- Rook, L. W. (1993). Laying down the law: Canons for drafting complex legislation. *Oregon Law Review*, 72, 663-688.
- Sadeghian, A., Sundaram, L., Wang, D. Z., Hamilton, W. F., Branting, K., & Pfeifer, C. (2018). Automated Semantic edge labeling over legal citation graphs. *Artificial Intelligence and Law*, 26, 127-144. <https://doi.org/10.1007/s10506-018-9217-1>
- Sannier, N., Adedjouma, M., Sabetzadeh, M., & Briand, L. (2015). An automated framework for detection and resolution of cross references in legal texts. *Requirements Engineering*, 20, 1-23. <https://doi.org/10.1007/s00766-015-0241-3>
- Sannier, N., Adedjouma, M., Sabetzadeh, M., & Briand, L. (2016). Automated classification of legal cross references based on semantic intent. In M. Daneva & O. Pastor (Eds.), *Lecture Notes in Computer Science: Vol. 9619. Requirements engineering: Foundation for software quality* (pp. 119-134). [https://doi.org/10.1007/978-3-319-30282-9\\_8](https://doi.org/10.1007/978-3-319-30282-9_8)
- Saw, K., & Sawyer, A. (2010). Complexity of New Zealand's income tax legislation: The final installment. *Australian Tax Journal*, 25, 213-245. Retrieved from <http://www.taxinstitute.com.au/publications/australian-tax-forum/australian-tax-forum>
- Spilker, B. C., Ayers, B. C., Barrick, J. A., Outslay, E., Robinson, J. R., Weaver, C. D., & Worship, R. G. (2019). *McGraw-Hill's essentials of federal taxation* (2019 ed.). New York, NY: McGraw-Hill Education
- Stevenson, D. (2014). Costs of codification. *University of Illinois Law Review*, 2014, 1130-1173.
- Surrey, S. S. (1969). Complexity and the Internal Revenue Code: The problem of the management of tax details. *Law and Contemporary Problems*, 34, 673-710.
- Tress, W. (2010). Lost laws: What we can't find in the United States Code. *Golden Gate University Law Review*, 40, 129-164.

Trosberg, A. (1997). *Rhetorical strategies in legal language: Discourse analysis of statutes and contracts*. Tübingen, Germany: Gunter Narr Verlag Tübingen.

Williams, C. (2007). *Linguistic Insights. Studies in Language and Communication. Vol. 20*. (2nd Rev ed.). Bern, Switzerland: Peter Lang.