

University of Denver

Digital Commons @ DU

Electronic Theses and Dissertations

Graduate Studies

3-1-2012

Development of a Measure Assessing Knowledge and Use of Internal Punctuation to Signal Syntactic Relationships

Pamela S. Van Horn Howard
University of Denver

Follow this and additional works at: <https://digitalcommons.du.edu/etd>



Part of the [Educational Assessment, Evaluation, and Research Commons](#), and the [Teacher Education and Professional Development Commons](#)

Recommended Citation

Howard, Pamela S. Van Horn, "Development of a Measure Assessing Knowledge and Use of Internal Punctuation to Signal Syntactic Relationships" (2012). *Electronic Theses and Dissertations*. 295.
<https://digitalcommons.du.edu/etd/295>

This Dissertation is brought to you for free and open access by the Graduate Studies at Digital Commons @ DU. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ DU. For more information, please contact jennifer.cox@du.edu, dig-commons@du.edu.

DEVELOPMENT OF A MEASURE ASSESSING KNOWLEDGE AND USE OF
INTERNAL PUNCTUATION TO SIGNAL SYNTACTIC RELATIONSHIPS

A Dissertation

Presented to

the Faculty of the Morgridge College of Education

University of Denver

In Partial Fulfillment

for the Requirements of the Degree

Doctor of Philosophy

by

Pamela S. Van Horn Howard

March 2012

Advisor: Dr. Kathy Green

©Copyright by Pamela Sue Howard 2012

All Rights Reserved

Author: Pamela S. Van Horn Howard

Title: DEVELOPMENT OF A MEASURE ASSESSING KNOWLEDGE AND USE OF INTERNAL PUNCTUATION TO SIGNAL SYNTACTIC RELATIONSHIPS

Advisors: Dr. Kathy Green

Degree Date: March 2012

ABSTRACT

A literary debate has been ongoing from the early 1900s regarding not only the place of grammar instruction in the classroom but even of the veracity of a grammatical standard, such as Standard American English, by which grammar skills may be measured. Very little empirical research has been attempted to compare the effectiveness of teaching methods because grammar assessment has been given even less attention. Therefore, to address this gap in the literature and to explore whether objective testing is suitable for assessing grammar skills, a 51-item measure was constructed to test the use of internal punctuation (commas, semicolons, and colons) and identification of syntactical structures (phrases and clauses). Rasch analysis found the measure as a whole possibly supporting a second dimension; therefore, the measure was analyzed as two scales: (a) a 34-item test of internal punctuation use and (b) a 16-item identification of syntactic structures. Both scales were found to be sufficiently unidimensional and reliable. In addition, scalar invariance of both was determined through DIF analysis by educational level. Validity evidence was obtained through a series of correlations with survey items assessing self-confidence and knowledge of the constructs tested in each scale. With the promising results of this endeavor in that objective testing can be effective, perhaps the debate may inspire researchers and educators alike to consider formal instruction of grammar in the context of a standard.

Acknowledgments

With great respect and gratitude, I would like to acknowledge the invaluable support and guidance of my dissertation committee members, Dr. Kathy Green, Dr. Duan Zhang, and Dr. Elaine Cheesman. Each was endlessly encouraging and patient. In addition, I owe thanks to Dr. Nick Cutforth, who also gave freely of his time and energy to see this project to the end. I am honored to have learned from all of you.

Table of Contents

Chapter I.....	1
Introduction.....	1
Statement of the Problem.....	2
Statement of Purpose.....	4
Literature Review.....	5
Broad Historical Framework of Grammar Teaching.....	5
Formal Grammar: Traditional versus Prescriptive.....	8
Transformational/Generative Grammar.....	12
Sentence Combining.....	15
Grammar Instruction in the Context of Student Writing.....	17
Functional or Rhetorical Grammar.....	18
Grammar Instruction Today.....	20
Attitudes toward Writing Instruction.....	21
Conclusion.....	23
Measurement of Grammar Perceptions and Skills.....	24
Measuring Language Skills.....	24
Construction of Objective Measures.....	28
Content.....	28
Format.....	29
Types of items in the multiple-choice test.....	30
Question stems.....	31
Content.....	31
Distractors.....	31
Number of choices.....	31
Writing Sample.....	31
Data Analytic Technique: Rasch Model.....	32
Research Questions.....	35
Correlations among Measures.....	36
Definitions.....	39
Delimitations.....	40
 Chapter II	 43
Method.....	43
Measure Development.....	43
Phase 1: Planning.....	44
Literature Review.....	44
Audience.....	44
Phase 2: Construction.....	44
Construction of the Multiple-Choice Grammar Test: SRSIP.....	45
Grammar Assessment Content.....	45
Grammar Assessment Format.....	45
Types of items in the multiple-choice test.....	45
Question stems.....	46

Content.....	46
Distractors.....	46
Number of choices.....	46
Related survey items.....	47
Phase 3: Quantitative Evaluation.....	47
Participants.....	47
Procedure.....	49
Item Analysis.....	50
Unidimensionality.....	50
Analysis of SRSIP.....	52
Phase 4: Validation.....	52
Writing Sample.....	52
Chapter III.....	54
Results.....	54
Rasch Analysis of All SRSIP Items.....	54
Dimensionality.....	54
Rasch Analysis of SRSIP Punctuation Items (Test Sections 1-3).....	56
Dimensionality.....	56
Reliability.....	60
Invariance.....	61
Targeting.....	62
Rasch Analysis of SRSIP Phrase/Clause Identification Items (Section 4).....	64
Dimensionality.....	64
Reliability.....	65
Invariance.....	67
Targeting.....	67
Validity Evidence for SRSIP Measures.....	68
Distributions.....	68
Bivariate correlations among validating survey items.....	68
Correlation between person logit scores for Punctuation Items and Phrase/Clause Identification Items.....	70
Correlations among person logit scores for Punctuation Items and survey items.....	70
Correlations among person logit scores for Phrase/Clause Identification Items (minus item 4-3) and survey items.....	71
Chapter IV.....	74
Discussion.....	74
Governing Theory for Construction of SRSIP.....	74
Evaluation of SRSIP Test Texts.....	76
Theory behind SRSIP Response Choice Sets.....	77
Dimensionality of SRSIP Scales: Punctuation Items and Phrase/Clause Identification Items.....	78
Reliability of SRSIP Scales and Sample Consideration.....	79

Validity of SRSIP Scales: Veracity of Writing Sample and Survey Items as Useful Validation tools.....	80
Limitations.....	80
Further Research.....	82
Exploring invariance.....	82
Tailoring item construction for specific samples.....	83
Continued quantitative analysis of the writing sample.....	84
Teaching methods.....	84
Validation of SRSIP.....	84
Conclusions and Implications for Practice.....	85
Implications.....	87
References.....	90

List of Tables

Table		
1	Self-Assessment Survey Items.....	36
2	Related Survey Items: SRSIP Multiple Choice Test.....	37
3	Writing Sample Rubric.....	38
4	Scale Development Procedure.....	42
5	Largest Standardized Residual Correlations for Punctuation Items.....	59
6	Correlations among Punctuation Items Person Logit Scores and Validating Survey tems.....	71
7	Correlations among Phrase/Clause Identification Items Person Logit Scores and Validating Survey Items.....	72

List of Figures

Figure

1	Standardized Residual Plot for Contrast 1 of Punctuation Items.....	58
2	Table of Standardized Residual Variance (in Eigenvalue units).....	60
3	Persons Map of Punctuation Items: Winsteps Output.....	63
4	Persons Map of Phrase/Clause Identification Items Minus Item 4-3: Winsteps Output. .	66

List of Appendices

Appendix A.....	100
Appendix B.....	102
Appendix C.....	104
Appendix D.....	124
Appendix E.....	125
Appendix F.....	126
Appendix G.....	127
Appendix H.....	128
Appendix I.....	129
Appendix J.....	130

Chapter I

Introduction

Standard American English. At one end of the spectrum, some vehemently defend its tenets for their beauty and order, can identify every structure in the language by name, and will correct even casual acquaintances on their grammatical blunders. At the other end, some decry the existence of Standard American English as myth, citing the evolution of language as determined by its speakers, evidenced in the predictable structures of English dialects; these skeptics note the multiple aberrations in the rules and arcane nomenclature of language structures and temper their homage to Standard American English by designating it merely as one of many English dialects. Speakers and writers of both extremes, however, will use Standard American English if their arguments are to be heard. That is the fact; only a few departures are tolerated for publication in any given prose genre: academia, journalism, even fiction in the name of style. However, students are graduating from secondary education with poor grammar skills. ACT (2004) reported that 46% of the 2002 graduating seniors taking the ACT Assessment scored at or below a score of 19 on the ACT English Test. This score suggests these students are not fully prepared for college-level coursework and may struggle with basic fundamental

English skills in punctuation, linking clauses effectively, and agreeing subjects with verbs. Amid these two arguments, that grammar use either does or does not significantly contribute to written communication, sit students in classrooms, receiving numerous amalgams of methods reflecting these extremes if they receive instruction in English grammar at all. An additional unknown is whether teachers across the curricula are confident in their own understanding of grammar and what effect a lack of confidence may pose in teaching grammar or holding students accountable for their grammar use.

Statement of the Problem

As will be discussed in the literature review, many notions of instruction in grammar exist in elementary and secondary education without consensus among educators and researchers as to the most effective and efficient means of teaching the subject. The research itself, owing to limitations of methodology, confounds the efforts of educators to choose teaching methods that help raise student writing performance to conform to those standards mandated by state testing requirements. No consensus of findings suggests clear guidance in choosing one methodology over another. In addition, owing to a lack of research in this area, neither traditional objective assessments of grammar nor traditional essay grading practices focusing on error have been shown to accurately measure students' understanding of grammar. Finally, even though instruction in grammar may be sporadic and inconsistent in its methodology, some students still learn Standard American English, perhaps owing to their own self-efficacy and/or to specific methods. In effect, student competencies have been measured with tests that presume to accurately measure learning, but so much of the presumed accuracy is

dependent on the competencies or theoretical leanings of the assessor, such as the grader, teacher, or the removed researcher. Isolated objective tests have not been shown to account for the many dynamics and processes governing students' decisions in grammar use, such as those associated with semantic, syntactic, and lexical choices (Petrovitz, 1997). In addition, measures of proficiency in grammar use are often created without the foresight needed to undergo rigorous analysis, such as those to determine invariance and unidimensionality, an assumption required for analysis using item response theory (IRT), the type of analysis currently at the forefront of measurement. For example, in their analysis of grammar placement tests using item response theory, Green and Weir (2004) explored the use of such tests to serve as diagnostic tools for instruction. They suggested that unidimensional test items must be able to be ranked along a continuum of difficulty and that the ability of the test takers be scored along the same scale with respect to their range of abilities and skill attainment. Their findings regarding the Global Placement Test (Hughes, Weir & Porter, 1995 as cited in Green & Weir, 2004) found that the difficulty of the grammar items was related more to test method than to the items' presumed linguistic difficulty. This example suggests that results of measurement of grammar proficiency, even when instruments are carefully constructed, are not always able to provide evidence of proficiency or may fail to measure it altogether. Here, the items themselves were not necessarily able to be ranked in difficulty because difficulty was a function of instructional method. Linguistic difficulty is a construct that is hard to establish in the first place because of the lack of research in this area. If many types of grammar items are not scalable in terms of their difficulty, then the effectiveness of such

items cannot be ascertained. Consider further then the types of tests likely to be composed or used by classroom teachers, who may not have a clear understanding as to what extent the tests they are using are measuring the grammatical constructs they may assume they are testing. Yet student proficiency is judged by such tests. In addition to the problems of measurement and inconsistent teaching methods is the question of whether teachers themselves, not just language arts teachers, are proficient in grammar as measured by objective tests and writing samples.

Statement of Purpose

Few English educators would probably argue with the sentiment expressed by students that grammar is a very difficult subject to master. This difficulty could be informed by several components associated with the grammar instruction that students have received. As discussed in the following literature review, both the time devoted to teaching grammar as well as the methods used vary widely from school district to district. Therefore, determiners are objective measures of participants' specific grammatical skills in terms of conformance to Standard American English (SAE). These two components inform one another and were measured in this study: (a) grammatical knowledge of specific aspects of grammar (i.e., knowledge of syntactical structures and use of internal punctuation to mediate syntactical structures) and (b) for the purposes of validation, perceptions of students' grammatical knowledge of syntactic structures and internal punctuation.

In addition, a writing sample, wherein participants discussed the importance and impact of their own knowledge and comfort with SAE, was obtained to gather additional

validating data on respondents' educational experiences regarding grammar instruction not captured in the quantitative instruments and to compare their conformity to SAE as assessed in a multiple choice test (i.e., use of commas, semicolons and colons as well as prevalence of comma splices, run on sentences, and fragmented sentences). It was expected that the writing sample would yield scores of aspects of grammar use in self-produced writing and to provide evidence of construct validity.

Literature Review

Broad Historical Framework of Grammar Teaching

The teaching of grammar has been fraught with much open debate. On one hand, native speakers intuitively understand the grammaticality of their language. One does not have to be directly taught rules of syntax, for instance, that such a sentence as “John going is to the store” is ungrammatical in English. But for some English speakers, “John be going to the store” is grammatical in his or her own English dialect even though such usage is considered unconventional and by some standards ungrammatical. As a consequence of English speakers' dialectal variants, much emphasis in teaching grammar has been placed on English usage, such as subject-verb agreement. Under the guise of *proper grammar*, the cadre of usage rules is often epitomized by somewhat illogical rules, such as neither ending sentences with prepositions nor splitting infinitives (what Hartwell, 1985, terms *linguistic etiquette*). Many of these rules continue to circumvent the quest to have all students conform their grammar usage to an agreed upon standard and are still tools of the linguistic hegemony to retain the division of dialectal speakers from the elite (Micciche, 2004).

English teachers have been caught up in this tendency at the expense of student learning. Yet detractors have always noted that if the rules are clear, then why are so many exceptions readily apparent even in the language of the purveyors of SAE? Many public arguments have focused on traditional methodology of teaching grammar. This methodology is comprised primarily of rote memorization of parts of speech and components of syntax, and learning is facilitated and assessed by exercises and tests using decontextualized language in sentence form. The memorization of the associated grammatical metalanguage, or the language of grammar, is the foremost goal of this methodology and for many decades was believed to be sufficient in improving student writing. But available published articles refuting this argument date back to the turn of the century. For instance, Rapeer's 1913 study replicated Hoyt's 1906 study (as cited in Rapeer, 1913) designed to test the leading arguments for the teaching of grammar, i.e., that it: (a) disciplines the mind; (b) prepares the way for the study of other languages; (c) gives command of an indispensable terminology; (d) enables one to use better English; and (e) aids in the interpretation of literature. Both Rapeer and Hoyt rejected outright the first three arguments, and focused on the final two, expecting to find sufficient correlations among students' knowledge of grammar and their skills in composition and literary interpretation if such arguments were true. No such correlations were found in either study when they compared students' skills in the areas of grammar, composition, and literature interpretation.

Rapeer's (1913) and Hoyt's (1906) arguments against such embedded notions of grammar taught as a separate subject outside the context of students' own writing,

especially when taught as such at the elementary level, are singled out because their arguments have been cycling through the literature ever since. The research community on the subject of teaching grammar is very familiar with the historical arguments and with the findings and limitations of the ongoing research, but very few will adamantly argue that one method works for most students. (For reviews of research on grammar instruction, see Andrews et al., 2006; Devet, 2002; Kolln & Hancock, 2005; Wyse, 2001, 2006.) Consequently, the place for grammar in the present-day curriculum is no more agreed upon than ever and has been largely ignored in the absence of definitive findings. In the meantime, teachers are frustrated with students' grammar skills, but they are no more prepared than are researchers to advocate for one methodology over another even though they are the ones most responsible to students themselves and are held accountable to such state standards as these from the Colorado Department of Education (1995, p. 9-10):

STANDARD 3: Students write and speak using conventional grammar, usage, sentence structure, punctuation, capitalization, and spelling....

RATIONALE:

Students need to know and be able to use Standard English. Proficiency in this standard plays an important role in how the writer or speaker is understood and perceived. All skills in this standard are reinforced and practiced at all grade levels and should be monitored by both the teacher and student to develop lifelong learning skills...

GRADES 5-8: As students in grades 5-8 extend their knowledge, what they know and are able to do includes

- identifying the parts of speech such as nouns, pronouns, verbs, adverbs, adjectives, conjunctions, prepositions, and interjections;
- using correct pronoun case, regular and irregular noun and verb forms, and subject-verb agreement involving comparisons in writing and speaking;
- using modifiers, homonyms, and homophones in writing and speaking;
- using simple, compound, complex, and compound/complex sentences in writing and speaking;

- punctuating and capitalizing titles and direct quotations, using possessives, and correct paragraphing in writing...

In agreement with Hoyt, Rapeer (1913) asserts,

Grammar as ordinarily taught in the elementary school is abstract, relatively meaningless and beyond the needs and reasoning abilities of children...it tends to retard rather than promote the natural development of the child, taking up his time and standing in the way of his progress toward a fair use of English in the few years of his school life (p. 126).

Yet in 2008, these state standards suggest some adherence to just the type of program against which Rapeer argued. What are teachers and students alike to do? First, a review of the major research findings and methodological approaches is in order to establish the framework of possible instructional strategies available for teachers' use.

Formal Grammar: Traditional versus Prescriptive

At the foundation of the discourse on teaching grammar lies the unstable and shifting notion of *formal grammar* itself, what constitutes formal grammar, and what methodologies might best be applied in the teaching of it. (For a discussion of uses of the term *grammar* as used in the literature, see Hartwell, 1985.) Until relatively recently, within the last twenty years or so, formal grammar has been inextricably linked to SAE. The existence of SAE as the proper form of English to which Americans should aspire in their writing and speaking was largely undisputed. Teaching the forms of SAE were the stuff of grammar texts published regularly by the likes of Warriner (1988). In what has now become referred to in the literature as *traditional methodology*, against which Rapeer argued, formal grammar was taught as a separate subject in the English curriculum with a relatively limited number of innovative strategies. Course content

relied heavily on teaching grammatical terms (e.g., *infinitive phrases*) and identifying adjacent structures in decontextualized language. Grammar books served not just as grammatical resources but as grammatical texts. The strategy of diagramming sentences could also be considered part of a traditional grammar curriculum as well. The term *prescriptive grammar* is applied to this methodology in that students are taught both the rules of the language and that its Latinate metalanguage is sufficient to describe its parts and functions. Brown (1996) captured the sentiment of teaching grammar prescriptively under his heading of “Misconceptions Often Held by Students”:

Proper grammar and speech aren't the norm but a form of elitism. This burgeoning thought, which must be rapidly dispelled, is that speaking and writing correct grammar somehow represent a form of aristocracy. The usage of nonstandard speech or ‘street language’ in any of my classes is quickly and strongly discouraged because it is wrong; it is not correct (¶1).

Not all advocates for traditional grammar still argue for its inclusion in the English curriculum quite as didactically as does Brown, but most echo the CSAP rationale presented earlier, primarily that teachers deprive students of lifelong opportunities if they are denied the exposure to the language that carries the greatest social capital in the culture, that is, SAE (Basset, 1981; Benjamin, 2006; Hoffman, 2006; Nunan, 2005; Vavra, 1996, 2003). Presently, many who teach a formal grammar systematically would argue that they are not teaching grammar prescriptively; rather, they teach grammar to help students discover the functionality of their language and therefore the range of language possibilities. Kolln (1981, 1983; Kolln & Hancock, 2003), who along with Corbett and Finkle (1992), and Quirk, Greenbaum, Leech, and Svartvik (1985), has written well respected grammar resources and texts (1984, 2006; Kolln &

Funk, 2006), and argues for the formal teaching of grammar yet eschews the practices of the prescriptivists who impose nonsensical rules and test-makers who construct tests measuring only superficial errors (Kolln & Hancock, 2003). Instead, Kolln regards systematic formal instruction in grammar as a useful forum for discussions about language. She refers to her methodology as both *functional* and *rhetorical* rather than *traditional*.

Kolln's present argument is disputed, however, in light of the lingering notions that formal grammar is de facto prescriptive by many interpreters of the body of research on formal grammar instruction. Prior to 1963, most of those who publicly refuted the effectiveness of teaching grammar formally, which could be taken to mean teaching traditionally or even prescriptively, *versus not teaching it*, could rely on the findings of only a few quantitative studies, such as Hoyt, (1906) and Rapeer (1913) but most importantly Harris (1962, as cited by Braddock, Lloyd-Jones, & Schoer, 1963). In his unpublished dissertation, Harris compared error rate within essays produced by two groups of students, one having received formal grammar instruction and a control group, over a two-year period. Harris found no differences in the error rate in participants' essays between the groups. The introduction of Harris's study to the research community in Braddock et al.'s comprehensive review of writing instruction included the following conclusion, which echoed Harris's own:

In view of the widespread agreement of research studies based upon many types of students and teachers, the conclusion can be stated in strong and unqualified terms: the teaching of formal grammar has a negligible or, because it usually displaces some instruction and practice in actual composition, even a harmful effect on the improvement of writing (p. 37).

Fifty years earlier, Rapeer (1913) had expressed a similar sentiment, cited earlier, primarily that teaching grammar formally simply wastes precious instructional time. The force of Braddock et al.'s (1963) caution of harm challenged educators' notion of grammar instruction and essentially gave English educators license to drop grammar instruction altogether for fear of the harm it might cause, leaving others to argue in the vacuum created for some kind of instructional techniques that would help students make better writing decisions at the level of the sentence and in publicly-expected accordance to conventional standards of English. Few could argue that no systematic instruction at the sentence level was not equally harmful for its eventual impact on students' participation at all levels of societal opportunity, most of which depend on one's use of language (Kolln & Hancock, 1983). Nevertheless, the traditional and prescriptive methods of grammar instruction as a subject unto itself became the pariah of English instruction and retained its association with formal grammar. In 1985, the National Council of Teachers of English [NCTE] issued the following position statement:

On Grammar Exercises to Teach Speaking and Writing

This resolution was prompted by the continuing use of repetitive grammar drills and exercises in the teaching of English in many schools. Proposers pointed out that ample evidence from 50 years of research has shown the teaching of grammar in isolation does not lead to improvement in students' speaking and writing, and that in fact, it hinders development of students' oral and written language. Be it therefore resolved, that the National Council of Teachers of English affirm the position that the use of isolated grammar and usage exercises not supported by theory and research is a deterrent to the improvement of students' speaking and writing and that, in order to improve both of these, class time at all levels must be devoted to opportunities for meaningful listening, speaking, reading, and writing; and that NCTE urge the discontinuance of testing practices that encourage the teaching of grammar rather than English language arts instruction.

Transformational/Generative Grammar

In the early 1960s, educational applications of transformational/generative grammar (see Chomsky, 1957) began to be considered in the context of grammar methodologies, especially for its superior descriptive possibilities over prescriptive methods. Even Braddock et al. (1963) qualified their conclusions of the Harris (1962) study in that they did not necessarily apply to instruction in transformation/generative grammar. For advocates of systematic grammar instruction, use of principles of transformational grammar was enticing.

Transformational/generative grammar examines how language is generated and how transformations of language structures are made to create other language structures, for example, simplistic transformations from active into passive voice or declarative into interrogative sentences. Owing to its academic linguistic roots and the precision of its descriptions of language, the study of transformational grammar generated even more grammatical terminology to account for the imprecision and gaps in the Latinate terminology of traditional grammar.

Applications in the secondary classroom presented even greater time-challenges for adequate instruction, however. Given the harm potentially caused by formal grammar instruction, advocates were hard pressed to justify such time allocations unless such instruction was shown to transfer into better student writing.

Further research was conducted to compare the relative effectiveness of transformational and traditional grammar instruction, but only the most influential studies are discussed here: Bateman and Zidonis (1966) and Elley, Barham, Lamb, and Wyllie

(1975). In an individualized randomized controlled two-year study regarding the relationship between sentence formation and composition writing using a generative approach, Bateman and Zidonis found greater effects in structural complexity and grammatical operations in well-formed sentences for the group receiving generative grammar instruction as compared to the control group which had received no formal grammar instruction. These effects were attributable only to the performance of 4 of the 50 students however. Bateman and Zidonis concluded tentatively that secondary students can learn generative grammar, which may help to reduce their errors in writing, a weak finding in light of the arguments against such formal instruction.

Surpassing Bateman and Zidonis's (1966) and studies prior (e.g., Harris, 1962) was the experimental rigor of Elley et al. (1975). In this study, three matched groups of middle-school aged students (neither of very high or very low ability; $n = 248$) were compared over a three-year period, one receiving traditional instruction, one receiving transformational instruction, and a control receiving reading and writing instruction without formal grammar instruction (i.e., rendered only on an ad hoc basis). This study used two prepared, named curricula for the traditional and transformational groups; the control groups used non-grammar components of one curriculum along with identified additional resource materials. Clustering effects were accounted for by rotation of teachers, who taught all groups in one or more of the sections of each grouping. Measures were administered on 12 variables, including essay composition, sentence combining, usage, and mechanics, as well as an attitudinal measure. As Elley et al. noted, with such experimental controls in place and sufficient time given to evaluate growth over time, if

significant differences were to be found, these conditions would best allow for their emergence. No overall group differences were found, however, except for some inconsistent effects between group pairings fleshed out by further analysis of item differences on individual tests. In other words, no group emerged as having improved considerably in their writing as an effect of any of the three types of instruction. In fact, the greatest difference found was in student attitudes toward instruction at the end of the experiment: students in the transformational group reported liking writing and reading less as a result of the instruction than did the other groups. Elley et al. concluded that formal grammar instruction, either traditional or transformational, simply provided no transferable benefits—nor harm—to students’ composition skills.

Andrews et al. (2006) pointed out three common problems with research regarding comparisons with formal methods and other methods: (a) as in the Harris study, specific descriptions of neither the methodologies nor the measures were adequately provided; (b) the clustered nature of the data was not explored for implications of effectiveness—in other words, teacher rapport and competence may have contributed to results, but the extent to which they did is unclear; and (c) insufficient time was accorded to study effects over time. Elley et al. (1975) designed their study to counter such criticism and found no effects whereas Bateman and Zidonis (1966), with less scientific rigor, found some differences in favor of transformational grammar instruction but cautioned against overzealous interpretation of the results as an endorsement of such instruction. These two studies proved most influential in moving English educators and researchers to explore methodologies other than formal grammar instruction.

Sentence Combining

Sentence-combining, one such methodology, seeks to integrate principles of transformational grammar into exercises wherein, without the focus on the accompanying metalanguage, students are encouraged to expand their repertoire of syntactic forms and sentence length. As described by Rose (1983), transformations occur in sentence combining when kernels of sentences either disappear or are embedded, subordinated, or coordinated within another sentence during the combining process. Although Rose recounts 100 years of sentence-combining exercises in educational materials, the methodology gained renewed interest with the publication of research by Hunt (1965), Mellon (1969, 1981), Hunt and O'Donnell (1970), O'Hare (1973), Combs (1976, 1977), and Daiker, Kerek, and Morenberg (1978). In drawing a significant distinction between traditional grammar instruction and sentence combining, Strong (1976) encouraged teachers to, "Spend your time in putting sentences together rather than in taking them apart. It's the combining that counts" (p. 60).

These researchers of sentence-combining (cited above) regarded higher numbers of T-units per sentence, to include numbers of words, words per clause, and clauses per T-unit, as the best indicators of syntactic maturity. A T-unit is defined as one main clause plus any subordinate clause or nonclausal structure that is attached or embedded in it (Hunt, 1970, p. 4). Faigley (1980) criticized Hunt's T-unit as the best indicator of syntactic maturity, given that a sentence by sentence unit count decontextualizes the text and ignores the value of sentence variety and length to serve authors' various purposes. Compare the syntactic needs of audiences reading either a set of instructions or a

philosophical discourse, for example. The audience is served well by a series of shorter sentences in the instructions in terms of retaining information whereas shorter sentences in philosophical discourse may interject emphases to impress an audience with the importance of the point being made. Crowhurst's (1983) review of research in sentence combining, including the above cited studies, corroborates Faigley's assertions in that rarely were essays containing sentences of greater T-unit length rated higher in overall writing quality than those of control groups. As Rose (1983) concluded, the force behind sentence combining was simply to help students move beyond the composition of simple sentences to increase their confidence in exploring new syntactic forms. Each of the studies cited above showed findings of significantly increased T-units in the participant writing samples, yet, as Faigley notes, the count of T-units alone failed to account for syntactic maturity, fluency, complexity, and growth in terms of writers' subjects, purposes, and audience. Marzano (1976) also criticized some of the research methodology used by the researchers cited, noting the non-random forced-choice rating system used by O'Hare of writing samples between control and experimental groups as a vehicle to skew results in favor of the experimental sentence combining group, concluding that sentence combining may have improved overall writing quality but only to a certain extent. Finally, Vavra (1996) further noted that longer sentences generated in the studies did not necessarily mean that the sentences were more correct according to standard conventions. Nevertheless, the research findings gave educators promise of some systematic instruction to replace formal grammar, and sentence-combining texts

(e.g., O'Hare, 1975; Mellon, 1969; Moffet 1968; Strong, 1973) have since occupied a corner of the market of English curricula as well a significant voice in the debate.

Grammar Instruction in the Context of Student Writing

Whereas expanded use of syntactic forms in student writing has been reported with use of instruction in sentence combining, this methodology still does not address error as does traditional or transformational grammar instruction, both of which focus on teaching the standard language conventions either prescriptively or descriptively. With the advent of process writing methodologies that teach strategies for prewriting, writing, and editing (see Elbow, 1986; Atwell, 1987), attention to grammatical errors was relegated to the final editing stage of composition. One important shift within process writing is that texts for grammar instruction moved from the decontextualized language of grammar books to the students' own language on the page. Weaver (1996a, 1996b) was instrumental in moving educators to address writing issues, especially grammar, in the context of student writing, using mini-lessons as the forum for instruction and guidance. The potential of such methods perhaps lies in the individualization of instruction itself, which conceivably can blend inductive lessons in error correction and syntactic flexibility as needed by each student. Such instruction also meets students at their respective developmental levels because their own language is addressed. Because of the individualization, time allocation to this methodology has been supported despite the challenges it poses to incorporate it systematically into the curriculum (Weaver, McNally, & Moerman, 2001). Research findings are nevertheless lacking to show effects of such instructional methods.

Functional or Rhetorical Grammar

One final consideration in tracing the historical trail of grammar instruction is the influx of voices calling for focus on functional or rhetorical grammar instruction. Kolln (1981) points out that all the methodologies in vogue, sentence combining included, de facto teach grammar in some form or another. Consequently, she advocates systematic instruction in the structure and conventions of the language primarily from a functional perspective to educate language learners as to the possible array and purposes of language use. In so doing, students can then consider the language uses they encounter, including their own, from a rhetorical perspective: to what audiences are writers appealing and what does their language use accomplish? Kolln (1984, 2006; Kolln & Funk, 2006) argued that if the structure and conventional uses are not well known to students, a rhetorical examination is unlikely to be useful. Fearn and Farnan (2007) tested their method of functional grammar instruction that focused on parts of speech. The authors deemed their program as one in which students studied grammar *in* writing rather than *for* writing. Students wrote practice sentences with given sets of requirements for the inclusion of certain parts of speech. Fearn and Farnan compared students receiving the instruction in directed writing with a control group receiving traditional grammar instruction and separate writing practice. In their assessments of approximately 150 tenth grade students, the experimental group showed gains in a holistic writing assessment although no differences were noted in error rate or writing fluency. Noting that student writing improves over time, Fearn and Farnan considered the gains after only five weeks of instruction to give some direction in valuable ways to incorporate systematic grammar

instruction into the curriculum. Such rhetorical examinations can move beyond such studies as Fearn and Farnun's study focusing on parts of speech. Both Dawkins (1995) and Petit (2003) advocated teaching punctuation in the context of the rhetorical implications of its use. Both asserted that use of punctuation is not rule-bound as is syntax in many ways, but punctuation can clarify syntactical relationships, signal subordination versus coordination, for example, and broaden the writer's repertoire of syntactical possibilities.

Others have expanded on this thinking and advocate grammar instruction in the context of its social force in meaning making. To these proponents of grammar instruction, SAE is considered the dialect of the elite, and they argue that many grammars exist within specific cultural constructs and settings. Micciche (2004) explained,

Rhetorical grammar analysis encourages students to view writing as a material social practice in which meaning is actively made, rather than passively relayed or effortlessly produced. The study of rhetorical grammar can demonstrate to students that language does purposeful, consequential work in the world—work that can be learned and applied (p. 716).

Schleppergrell (1998) described such a process of teaching rhetorical grammar use by breaking down language events by *genre*, or passages written in specific cultural contexts, and the associated *registers*, or lexical and grammatical features that realize the genre in terms of three metafunctions: ideational, interpersonal, and textual.

Schleppergrell argued that such focused examinations of different genres and the registers required to achieve their desired effect enables students to understand that grammatical choices are linked to specific contexts. Micciche regarded instruction in rhetorical grammar as a forum in which to teach critical thinking and cultural critique

rather than a set of static conventions to which few adhere. Schleppergrell also noted that research is needed to verify this argument. As early as 1974, the Conference on College Composition and Communication issued this forceful resolution:

We affirm the students' right to their own patterns and varieties of language -- the dialects of their nurture or whatever dialects in which they find their own identity and style. Language scholars long ago denied that the myth of a standard American dialect has any validity. The claim that any one dialect is unacceptable amounts to an attempt of one social group to exert its dominance over another. Such a claim leads to false advice for speakers and writers, and immoral advice for humans. A nation proud of its diverse heritage and its cultural and racial variety will preserve its heritage of dialects. We affirm strongly that teachers must have the experiences and training that will enable them to respect diversity and uphold the right of students to their own language.

Grammar Instruction Today

Twenty years later in 1994, in light of the above reasoning and in response to the disarray created by its earlier declaration that formal grammar instruction was deemed of little value, the NCTE issued a second proposal:

On Language Study

The teachers who proposed this resolution said the response of many teachers to the grammar debate has been either to avoid explicit instruction in the structure of English or to continue to teach grammar in a prescriptive manner. These extremes, they said, emphasize the need for NCTE to articulate strategies for developing the language awareness of teachers and students. Be it therefore resolved, that the National Council of Teachers of English appoint a committee or task force to explore effective ways of integrating language awareness into classroom instruction and teacher preparation programs, review current practices and materials relating to language awareness, and prepare new materials for possible publication by NCTE. Language awareness includes examining how language varies in a range of social and cultural settings; examining how people's attitudes vary towards language across culture, class, gender, and generation; examining how oral and written language affects listeners and readers; examining how "correctness" in language reflects social-political-economic values; examining how the structure of language works from a descriptive perspective; and examining how first and second languages are acquired.

This review reflects the state of grammar instruction today: no single methodology is endorsed and teachers are encouraged to continue exploring possibilities for integrating grammar instruction into the writing curriculum. With no consensus among educators, it can be inferred that the grammar instruction students are receiving in the United States is on the whole fractionalized and possibly sporadic. Ironically, given time frame between the move away from teaching grammar formally, especially prescriptively, to the inconsistent instruction today, current teachers, especially younger teachers, also have been taught grammar under these conditions. This state alone may account for much of the inconsistency. If teachers are neither knowledgeable of SAE (much less the more nuanced approaches to grammar, such as grammar as a cultural construction), nor confident in their own grammar skills, then student outcomes are tied to those of their instructors, not just language arts teachers but all teachers who require written work from their students.

Attitudes toward Writing Instruction

Few studies have attempted to survey students and teachers on their attitudes toward their grammar instruction. In their action research study of second, fourth, and eighth grade students, Hutchinson, McCavitt, Rude, and Vallow (2002) surveyed teachers across the curriculum, parents, and students regarding grammar instruction and support for grammar instruction. Student groups in the study performed at or above the 50th percentile in their state achievement tests. Both teachers and parents reported teaching and/or supporting grammar instruction, yet student responses indicated frustration and confusion with such instruction even though they highly endorsed grammar instruction

with 58% reporting that grammar instruction improved their writing. Hutchinson et al. did not report the methods used by teachers in the participating schools or the actual frequency of grammar instruction. For example, whereas teachers reported teaching grammar in their respective disciplines, correcting mistakes in oral and written communication as well as expressing confidence in their own abilities to teach grammar and in the transference of grammar instruction into writing, students reported a different picture: they proofread their writing for grammar primarily just when asked to do so by teachers, and they did not consider grammar correctness particularly important in classes other than English. Because the actual grammar instruction in the participating schools was not documented in the study, in their discussion of the implications of their findings, Hutchinson et al. did not attempt to reconcile these differences by focusing on the use of student responses to inform improved practices. Rather, they discussed possible reasons and strategies to improve practice as discussed in the literature. With such a paucity of research in this area, further research is warranted.

In the pursuit to find a sample that most competently could provide data to be used in assessing newly constructed grammatical measures, a few choices present themselves. Whereas high school students may offer rich data in assessing their grammar instruction, because of their widely varying language competencies, the data may be unreliable. College students would likely be better prepared to provide more reliable data as they are asked to write formally on a fairly consistent basis and may have greater cognizance as to how their previous instruction served them as they receive grades for their writing. However, this choice of group is biased in so far as non-college students are

not asked to participate. Therefore, it is possible that teachers, especially those pursuing their own education, could provide reliable data in the assessment of grammatical measures. As educators in the era of state-mandated testing, they are tasked with guiding students to meet state benchmarks for student performance, including proficiency in grammar. As they are in daily contact with student use of grammar and as they are confronted with their use of grammar in their own studies, their cognizance of grammatical issues may be more acute than that of the other groups identified.

Conclusion

Educators and researchers alike report varied theoretical underpinnings for modes of grammar instruction, yet it is unknown the degree to which grammar instruction occurs. Especially for college bound students, instruction in grammar is likely important to their post-secondary writing success. In the 2004 ACT National Curriculum survey, ACT reported that high school English teachers and college instructors differed most on one point: students' grammar skills. College instructors counted them as most important whereas high school teachers considered them least important. Of the high school teachers responding to the survey, 69% reported teaching grammar, 90% sentence structure, 83% punctuation, and 92% style, yet they devoted more effort to teaching writing strategy (96%) and organization (92%). Many would consider sentence structure, punctuation, and style under the purview of grammar instruction, so it is unclear the nature of the instruction being reported under *grammar*. This discrepancy is evidence of the widely varying views of grammar instruction that will require ongoing research to

parse. In the meantime, college instructors, employers, and the public at large are expecting high school and college students to graduate with effective grammar skills.

Measurement of Grammar Perceptions and Skills

Measuring Language Skills

In much the same way that the debate on teaching grammar has yielded no conclusive answers, the debate in language assessment in general and assessing grammatical skills in particular is equally inconclusive. Grammar skills traditionally have been assessed two ways: either by using (a) objective tests comprised of sentences of decontextualized language or paragraphs containing grammar errors to be corrected or (b) by grading students' writing samples for grammatical errors. In other words, error is generally the focus, and the error is assessed at the level of the sentence, a focus that often precludes the assessment that contextual lexical choices often drive grammatical decisions from one sentence to the next (Petrovitz, 1997). Ultimately, error counts are made as the writing deviates from SAE and presuppose a philosophical preference for the notion of a standard language, which does not attend to notions of grammar instruction in light of its uses in cultural contexts. To date, research is lacking in the assessment of grammar as used in specific cultural contexts in academic settings as described by Micciche (2004). In addition, research in the effectiveness of individualized instruction as advocated by Weaver (1996a, 1996b) are also absent from the literature.

The debate in language assessment is dichotomous. In one camp are gathered those who call for authentic assessments and in the other are those who advocate objective performance testing (see Alderson & Hughes, 1981). Morrow (1981) points out

that that to test language in any form, tests, authentic or objective, must break down language into discrete constructs, but he argues that every facet of language is interwoven. He argues that simply because one may test for components of language, the assessments do not reveal one's knowledge of the language. For example, whereas this idea of interdependency may seem apparent when attempting to test reading comprehension, less obvious is the inextricability when testing language structures, such as knowledge of phrases and clauses or use of commas; nevertheless, one could argue that such testing is equally problematic because, as Morrow asserts, it is unknown the extent to which such testing translates across all settings in which language is produced and comprehended. Consequently, the assessments, even if evidence for their face and content validity exists, still may be deemed unreliable and their constructs impossible to validate. Members in both camps tended to agree on this point (Alderson 1981a, 1981b; Moller, 1981; Morrow, 1981; Weir, 1981).

Morrow (1981) criticized objective tests because (a) test takers do not produce language-- rather, they recognize forms; and (b) the language of the tests is the language of the test developer; therefore, what is revealed are the differences in language norms between the test taker and developer. Some additional objective testing strategies have been found wanting. Recall that counts of T-units were criticized for their narrow focus in sentence combining studies (Faigley, 1980). Authentic tests, such as writing samples or essay tests, on the other hand, ask test takers to produce language, which should allow them to demonstrate their actual communicative ability. Alderson (1981b) countered the notion of the primacy of authentic tests over objective assessments by asserting (a) that

the idea of authenticity in testing is essentially oxymoronic because the only thing authentic in a testing event is that it is a testing event, that the act of testing may change participants' use of language; and (b) if constructs cannot be validated in any kind of language testing, authentic or objective, then to assume that authentic testing is somehow better is baseless—no compelling body of evidence has been offered to negate the possible merits of testing either way, and no corpus exists to support either.

The arguments of both ends of the assessment debate have their merits, but the common ground is troublesome and deflating: construct validity remains elusive despite the form of testing, and therefore reliability is debatable. So what are educators to do? Are they to give up testing even if the testing is imperfect? Surely, some indicators of proficiency are revealed if test developers are conscientious in the construction process, especially to the language development of the test takers. One assessment strategy is to use both forms of testing, authentic and objective, in tandem. In his review of the literature of writing assessments, Cooper (1980) found agreement among authors and researchers of the time that essay tests and objective tests can be highly correlated if sufficient inter-rater reliability can be established for the scoring of essays. In conjunction, Cooper asserted that using both forms is the optimal strategy for assessing writing ability; however, for large samples, using both assessments is difficult and costly. Halpin, Halpin, and Schaer (1981) studied the correlations between holistically scored essays and the Missouri English Test and found that both forms of tests assess similar skills, yet the use of one without the other is insufficient. The objective test accounted for 26% of the variance in writing ability. They suggested that holistic scoring of essays

allowed raters to examine a broad range of abilities whereas objective measures focused on just a few. On the other hand, Michael and Shaffer (1979) compared the Test of Standard Written English (TSWE) and the California State University English Placement Test (Csuc-Ept) to gauge their prediction of grades in a basic English composition course and overall freshman-year grade point average. They found that the TSWE was equally predictive as the essay portion of Csuc-Ept of both domains. Considering all the objective tests and subtests together, Michael and Schaffer concluded that they superseded the predictability of either of the essay tests, and the objective tests required less than half the time to complete. The validity of a test is ultimately determined by its use. In this case predicting academic success was the purpose for administering these measures.

Perhaps owing to the general acceptance of SAE as the dominant dialect in America, assessing grammar in either form is possibly the most manageable of language assessments--for better or for worse. Inherent in testing conformance to SAE is the tacit belief that use of SAE is beneficial, useful, and necessary for test takers. If test takers do not believe it is any of those things, then reliability is impacted, and thus validation is even more elusive. Given the inconsistency with which grammar has been taught, it would be surprising if test takers did value a grammar assessment. Nevertheless, there are those who do value SAE, and those people are most likely to provide data that are useful in examining a measure for reliability. In addition, using both authentic and objective measures together may provide validity evidence. To provide additional important evidence, test takers themselves may offer valuable insight into the effectiveness of objective tests as they reflect respondents' understanding of their own use of grammar.

Consequently, a confidence item could provide additional insight into participants' decision-making processes. A self-assessment measure wherein respondents are asked to assess their knowledge of the skills tested by objective and authentic measures may also provide additional validity evidence.

Construction of Objective Measures

The construction of the objective measure requires adherence to the following principles of content and format.

Content

Because grammar is defined so broadly, it is most useful to narrow the scope of the objective tests to a set of grammatical principles widely agreed upon for their propensity to reveal core understanding of language structure, such as syntactical relationships (as described in transformational/generative grammar instruction and as practiced in sentence combining exercises) as well as the role that punctuation plays in signaling those relationships. For example, theorizing within the context of Functional Discourse Grammar, Hannay and Kroon (2005) described the role of punctuation to signal syntactic relationships: conceptually, writers create *discourse acts* that are strategic steps to either convey ideas or to regulate flow of either information or interactions of various discrete acts or both. They argued that punctuation functions to regulate these *moves* in the discourse among syntactic units. In terms of testing students' understanding of syntactical relationships, assessing punctuation use as a regulator of these relationships provides a manageable focus. This scope must be restricted to only a few predictable uses of punctuation likely to have been the subject of instruction and/or error correction

because punctuation use is often utilized as a vehicle of style. In other words, only a few principles of punctuation use must be tested in order to provide a valid quantifiable framework for assessment, and those principles must relate directly to the regulation of syntactic relationships. These points of punctuation are the comma, semicolon, and colon. Whereas dashes and ellipses are also used to regulate discourse moves within the sentence, these are less likely to be taught consistently because they are points of punctuation often used stylistically and are not emphasized to the same degree as the comma, semicolon, and colon among grammarians (see Corbett & Finkle, 1992; Kolln, 1984, 2006; Kolln & Funk, 2006; Quirk et al., 1985).

Format

Quantitative assessment of large samples necessary for gauging reliability and validity of measures requires careful control of item development and test format. Hambleton and Murphy (1992) discussed the criticisms leveled at objective testing in that it is essentially inauthentic because it fosters a one-right-answer mentality, narrows the curriculum, focuses on discrete skills, and underrepresents the performance of lower SES students (p. 4). Authentic measurement, or *performance testing*, on the other hand, resembles learning tasks and tests higher-order thinking skills, such as problem solving and critical thinking. Poorly constructed objective tests do not mirror instruction and rely primarily on information recall. Hambleton and Murphy noted that large-scale performance testing is feasible only if test constructors are trained adequately and if test conditions allow the time required for construction, administration, and scoring. They countered the criticism that objective tests, especially multiple choice formats, are unable

to capture learning, especially if one answer does exist, if the test items are necessarily and appropriately narrow and discrete, if the language of test items is culturally familiar to test takers, and if higher-order skills are assessed. David (2007) noted the criticism leveled against the use of multiple-choice format to assess grammar, that such testing narrows the following: grammar syntax and grammar use in semantic context. Additional criticisms asserted that poorly-written items may include uses debatable among grammarians.

When appropriately chosen as a testing format, multiple-choice offers two important advantages: ease of administration and scoring. However, construction of good multiple choice tests requires adherence to certain guidelines. In their review of 27 education measurement textbooks and 27 research studies and reviews, Haladyna, Downing, and Rodriguez (2002; see also Haladyna & Downing, 1989a, 1989b) created a taxonomy of 31 “item-writing rules” for the construction of multiple choice tests geared for specific multiple choice formats (p. 312). These rules cover item content, formatting, style, stem-writing, and choice writing. Pertinent to the measures developed for this study is the conventional multiple choice format comprised of a stem and a list of choices, recommended by all textbooks reviewed by Haladyna et al. Rules of particular interest are discussed below.

Types of items in the multiple-choice test. Haladyna et al. (2002) recommended that either a best or correct answer format be used in addition to avoiding a complex multiple choice format in which a test taker must supply more than one answer. They also recommended that items be phrased to minimize reading time.

Question stems. Two forms of stems figure most prominently in the literature: complete questions and stem completion. Whereas Haladyna et al (2002) found no differences in difficulty, they favored the complete question because it more clearly reveals the central idea of the test item. Directions should be clearly written so that test takers clearly understand the task.

Content. Haladyna et al. recommended that each item specify both content and a mental behavior. Items should be based on one important educational objective, avoid bias, and avoid cueing and trickery.

Distractors. Haladyna et al. suggested that distractors be plausible, logical, and reflect common errors associated with content.

Number of choices. In his meta-analysis of 48 articles, Rodriguez (2005) found that fewer choices reduce item difficulty and that 5 or more choices encourage guessing. He concluded that 3 or 4 choices with only plausible distractors included offered the greatest reliability.

Writing Sample

In contrast to objective measures, writing samples provide contextualized and relatively unbiased examples of participants' use of grammar and specifically their use of internal punctuation to mediate their own syntactical discourse. These uses of punctuation may provide interesting correlates to both the objective and self knowledge measures. Scoring comprised either of total error counts or error/observation percentages for items assessing overall quality of grammatical use, specific uses of punctuation, and prevalence

of comma splices, run on sentences, and fragmented sentences allow data that may be analyzed to provide validity evidence with items on the objective test.

Data Analytic Technique: Rasch Model

The item response theory (IRT) Rasch model (Rasch, 1960) allows for modeling of item difficulty as well as person ability. Rasch models, also known as latent trait models, assume that examinee performance is explained by an underlying latent trait. Therefore, an examinee's pattern of responses, rather than total raw scores on a test, accounts for ability. Rasch models do not assume that person ability and item difficulty are linearly related to the construct. Instead, IRT software producing an s-shaped item characteristic curve shows the relationship between trait level and probability of a correct response on any given item. Item difficulty is calculated from the number of persons in a suitable sample who endorse or succeed on the item, and the discrete responses, rather than the total responses, are considered manifestations of a latent trait as graphically depicted in the item characteristic curves. Rasch modeling differentiates between person ability and item difficulty, but both ability and difficulty are measured on a common metric. If the ability of an examinee is high, then the probability of his or her success on more difficult items is greater than that of those with less ability. Each examinee is expected to progress along the continuum of difficulty until items become too difficult to either answer correctly or endorse. Where item assessment is concerned, items are determined to be stable if the pattern of examinee success in terms of ability is stable. This stability is unlikely to occur if sets of items are not unidimensional, that is, more than one dominant trait is measured by a set of items. Therefore, it is critical that the

assumption of unidimensionality is met. Additional assumptions of all Rasch models are that the test responses on any two items are statistically independent and that the test is not governed by time constraints.

Item and scale invariance are also crucial features of Rasch modeling. Item invariance assumes that items will perform similarly across measurement events and samples. Invariance is conceptualized in four ways: conceptual/functional, operational, item, and scalar (Bond & Fox, 2007). For the measures developed in this study, conceptual/factual invariance regarding grammar understanding will most likely be unstable until an instructional mode for teaching grammar is widely endorsed and practiced with some consistency over time among heterogeneous samples and/or SAE realizes some significant shifts in acceptable use. Neither is likely to happen soon. However, operationalization of the constructs under study is presumed to be invariant unless shown to fail invariance. Item invariance (that the items will retain their meaning across measurements) is assumed unless shown to fail. If no differential item functioning is identified, constructs will be assumed to be invariant, at least for this sample.

Because probabilities are estimated, model fit is important in use of the Rasch model. The Rasch model software provides estimates of the parameters of item difficulty and person ability. Some research questions applicable to multifaceted designs analyzed by Rasch modeling may suggest the items will behave differently among groups of respondents, leading to differential item functioning, of the existence of distinct latent classes. The addition of an item discrimination parameter constitutes the 2-parameter model. Others advocate the estimation of an additional parameter, a guessing parameter,

which is known as the 3-parameter model. The addition of this parameter accounts for the possibility that person ability is partially dependent on the examinee's skill in guessing.

The choice of IRT model depends on the research question and the researcher's assumptions about the nature of measurement. If a compelling case can be made that items are likely to differ significantly in discrimination, then adding the item discrimination parameter is possible. If a similar case can be made for the addition of the guessing parameter, then the 3-parameter model can be used. However, most proponents of the Rasch model would disagree that these additional parameter estimations improve the model at all. As described by Hawkins (1987), item difficulty is the only parameter that actually can be estimated with any consistency or reliability. So even though Rasch proponents acknowledged item discrimination in all tests as well as the possibility of guessing, their assumptions are that discrimination is invariant and guessing does not occur or is very minimal. Their reasoning is that if tests are carefully constructed, then item construction will ferret out the tendencies for discrimination and guessing. Hawkins also points out that Rasch modeling creates a statistic based on a person's score that is wholly suitable to estimate that person's ability; nothing else is needed and, in fact, the guessing parameter does not generate a statistic at all to inform guessing. In short, the Rasch model is viewed as elegant and complete in itself as the model to ensure that measurement is true measurement. Nevertheless, studies comparing fit of the various models find very mixed results (Hambleton & Swaminathan, 1985). Often item discrimination indices do differ and guessing may occur unless the test is very easy. Nevertheless, practically speaking, additional parameters require a larger sample size and

more items for the estimates to be reliable, which make the Rasch model more convenient and most useful for many researchers. For the purposes of this study, the Rasch model was deemed appropriate to analyze the data.

Research Questions

This study was directed by the following research questions regarding *Syntactical Relationships as Signaled by Internal Punctuation: Multiple-Choice Grammar Test* (SRSIP)

- (a) Is a 51-item test measuring use of internal punctuation to regulate discourse among syntactic structures and identification of major syntactic structures unidimensional and is adequate reliability achieved for its use with a sample of adults with varying exposure to writing in their professional and /or academic duties as well as in their personal lives?
- (b) Are three multiple choice item sections measuring student understanding of internal punctuation to regulate discourse among syntactic structures collectively unidimensional and is adequate reliability achieved for their use with a sample of adults with varying exposure to writing in their professional and /or academic duties as well as in their personal lives?
- (c) Is a set of multiple choice items measuring student identification of major syntactic structures unidimensional and is adequate reliability achieved for its use with a sample of adults with varying exposure to writing in their professional and /or academic duties as well as in their personal lives?

(d) Does a series of survey items assessing knowledge of syntactical structures and confidence in use of internal punctuation and knowledge of syntactic structures positively correlate with person logit scores to sufficiently offer validity evidence of the measure (see Correlations among Measures below)?

Correlations among Measures

For the purpose of validating SRSIP, the following relationships within and among measure subscales and related survey items will be analyzed for statistically significant positive correlations:

(a) Various correlations among the SRSIP subscales’ person logit scores and sentence structure survey items. See Table 1 for items. Positive correlations are expected.

Table 1
Self-Assessment Survey Items

Item Construct	Items	Response Choices
Knowledge of Syntactical Terminology	1. I can recognize an independent, or main, clause	Rarely
	2. I can recognize a dependent, or subordinate, clause	Sometimes
	3. I can recognize a phrase	About half the time
	4. I can recognize a fragmented sentence	Most of the time
	5. I can recognize a run on sentence.	Almost Always
	6. I can recognize a comma splice.	
	7. I know how to punctuate between dependent and independent clauses	
	8. I know how to punctuate between two independent clauses.	
	9. I know when to use commas.	
	10. I know when to use semicolons.	
	11. I know when to use colons.	
	12. I understand the concept of modification within the context of sentence construction.	
	13. I understand the concept of subordination within the context of sentence construction.	

Note: *Reverse scored

(b) Various correlations among the two SRSIP subscales' person logit scores and related survey items regarding confidence in knowledge of sentence structures and using internal punctuation (see Table 2). Positive correlations are expected.

Table 2
Related Survey Items: SRSIP Multiple Choice Test

Item Group	Items	Response Choices
Confidence in Task Completion	How confident are you in your use of commas?	Not at all confident
	How confident are you in your use of semicolons?	Somewhat confident
	How confident are you in your use of colons?	
	How confident are you in your knowledge of phrases?	Confident
	How confident are you in your knowledge of clauses?	
		Very confident

(c) The SRSIP subscales' person logit scores and ratio of total error counts of comma, semicolon, and colon misuse and the prevalence of comma splices, run on sentences, and fragmented sentences to number of sentences within the writing sample. Negative correlations are expected (see Table 3).

Table 3
Writing Sample Rubric

Type of Punctuation or Error Type	Grammar Rules	Record Number of Sentences in Writing Sample	Record Error Count
Comma Use	1. Rule: When two independent clauses are joined by a coordinating conjunction, separate with comma.	Yes	Yes
	2. Rule: When an introductory dependent clause precedes independent clause, separate with comma.	Yes	Yes
	3. Rule: When an independent clause is followed by a dependent clause, no comma is needed to separate the two.	Yes	Yes
	4. Rule: Use a comma to set off an introductory word or phrase.	Yes	Yes
	5. Rule: Use commas to set off an appositive or parenthetical word or phrase.	Yes	Yes
	6. Rule: Use commas set off a nonrestrictive relative clause.	Yes	Yes
	7. Rule: No commas are needed to set off a restrictive relative clause.	Yes	Yes
Comma splices, run ons, and fragments	1. Rule: A run on sentence is comprised of two or more independent clauses adjacent to one another without proper punctuation; separate with a semicolon or period.	Yes	Yes
	2. Rule: A comma splice is comprised of two or more independent clauses adjacent to one another separated by a comma; use a semicolon or period instead.	Yes	Yes
	3. Rule: No internal punctuation is required for a simple sentence or clause with a compound verb wherein no clause or phrase interactions requiring punctuation exist.	Yes	Yes
	4. Rule: A fragment is a unit of words presented as a complete sentence but lacks an independent subject-verb relationship. (Error count here ignores interjections or expletives.)	Yes	Yes
Colon Use	1. Rule: No colon is needed if the structures preceding a list cannot stand alone OR a colon is needed if structures introducing items in a list can stand alone.	Yes	Yes

Note: Percentages are calculated for items for which number of sentences and error counts are recorded.
 Note: Totals are calculated for items for which error counts only are recorded.

Definitions

Formal Grammar—An instructional methodology wherein grammar is taught as a separate subject using decontextualized language as the texts of grammar instruction. Often called Traditional Grammar.

Functional or Rhetorical Grammar—Systematic instruction in the structure and conventions of the language primarily from a functional perspective to educate language learners as to the possible array and purposes of language use.

Grammar Instruction in the Context of Student Writing—Instruction designed to address grammatical issues in the context of student writing using mini-lessons as the forum for instruction and guidance.

Item Invariance—Provided that a measurement instrument contains items that have the same meaning across groups, items maintain their relative and absolute position on the latent trait.

Item Response Theory—A body of related psychometric theories that provide a foundation for scaling persons and items based on responses to assessment items.

Linguistic Etiquette—The socially sanctioned rules of usage in Standard American English.

Prescriptive Grammar—An instructional methodology of formal grammar instruction that emphasizes the rules of grammar as correct, unchanging, and sufficient in their description of Standard American English.

Sentence Combining—Seeks to integrate principles of transformational grammar into exercises wherein, without the focus on the accompanying metalanguage, students are encouraged to expand their repertoire of syntactic forms and sentence length.

Standard American English—A set of rules and descriptions of American English presumed to encompass the standard or correct forms of the language. Many argue that it is merely a dialect of the elite.

Transformational/Generative Grammar—Examines how language is generated and how transformations of language structures are made to create other language structures, for example, simplistic transformations from active into passive voice or declarative into interrogative sentences.

Delimitations

The primary delimitation to this study is with regard to the sample (see Participants under Methods). Anecdotally speaking, grammar is one of those topics that people either love or hate; few seem to have ambivalent reactions to it. A population of varied grammar skills, especially those tested by the measure under study, is warranted. However, such a population is not likely to volunteer unless required to do so, such as under such circumstances as fulfilling a course requirement, even though such a sample would be ideal: college students who are in a setting where writing is required by their programs and, hence, a heightened sense of their own grammar use is likely present, and their skills will vary, especially if students are recruited from many types of institutions of higher learning (e.g., public universities, private colleges, community colleges, and technical colleges). Owing to constraints in data collection, namely feasibility in

recruiting participation of instructors in institutions of higher education, this sample was not accessible. In order to recruit a sizable sample (over 300), which was desired for this project, the decision was made to make a broad appeal for participation among adults with varied writing needs. It was anticipated that this invitation was likely to be answered by those who are interested in grammar, perhaps pride themselves on their grammar use or have appreciation for the grammar instruction they received along their educational spectrum. This possibility proved true. A sample was recruited that was largely well-educated and, if not currently in school, considered writing to be an integral part of their professional responsibilities. This characteristic was favorable in that participants over all had a heightened sense of their grammar use. However, this participant set provided a limited range of responses to test items. Given that this test was designed to assess particular grammatical skills among those with a wide range of abilities, the decision to recruit the sample of respondents upon whom the analyses relied limited the generalizability of the measures' usefulness.

An additional limitation of the study is the use of item response theory as the sole theory used in analysis of scale reliability and unidimensionality. Factor structure may be best analyzed through as many theoretical bases as is feasible. In the case of the measure analyzed in this study, considering its unique construction, the decision was made that item response theory was sufficient for its analysis. Other methods of analysis, such as exploratory factor analysis or its superior, confirmatory factor analysis, which makes use of a variety of estimation processes, including maximum least likelihood, are useful in guiding the interpretation of SRSIP's value as an objective measure for assessing internal

punctuation use to signal syntactic relationships given larger sample sizes of greater variability and longer measures. As revealed in the literature review of this study, most practitioners of grammar assessment have not carefully attended to even the constructs assessed in their tests much less the test's format. The measure created for this study seeks to assess certain grammar skills using contextualized language (e.g., short anecdotes) and tailored multiple choice response sets. Therefore, item deletion per se is not feasible given the short tests although item revision is. Whereas another type of analysis could give additional insights into problematic item performance, IRT in its varied output was deemed to allow sufficient guidance in assessing item performance of this newly created measure.

Measurement error always poses threats to both reliability and validity. Should the measure developed for this study perform unreliably and/or if the measure cannot be considered valid, further analysis of the data would produce unreliable estimates.

Chapter II

Method

Measure Development

Four phases governed the development and validation of the *Syntactical Relationships as Signaled by Internal Punctuation: Multiple-Choice Grammar Test* (SRSIP). They were Phase 1: Planning; Phase 2: Construction; Phase 3: Quantitative Evaluation; and Phase 4: Validation (see Table 4).

Table 4
Scale Development Procedure

Development Phase	Scale Development Steps
Planning	Determine purpose of measure Conduct literature review Identify potential audience Identify potential participants Discuss potential procedures Select item format
Construction	Generate item pool Expert review for content validation Conduct cognitive interviews
Quantitative Evaluation	Field administration: Determine subscales Compare subscales to original purpose of instrument and revise Assess internal consistency reliability of subscales Assess participants' use of response format Optimize scale length and format
Validation	Assess convergent, and criterion-related validity by conducting correlational analysis among test scales and related survey items Assess relationship between expert and respondent interpretation of items by means of difficulty ratings

Phase 1: Planning

The planning stage was comprised of determining the purpose, scope, audience, and participants for the measures to be created for this study.

Literature Review. The literature review revealed pertinent themes in relation to the purpose: (a) types of theoretical grounding driving possible methodologies used to teach grammar, (b) types of methodologies used to teach grammar, (c) types of core grammatical concepts and skills students are expected to learn in elementary and secondary education, (d) guidelines in the construction of multiple-choice tests pertinent to grammar assessment, and (e) determination of the most appropriate data analysis to be used.

Audience. The audiences for results of this study were determined to be language arts educators and researchers seeking insight into the effects of current practice in the teaching of grammar.

Phase 2: Construction

Upon completion of the literature review, an item pool was developed for the measure. Items reflected the domains researched in the literature review (see Phase 1). Consultation with grammar experts was conducted to determine the clarity, propriety, difficulty of items, and reflection of the constructs of the items to reveal respondents' use of grammar. In addition, cognitive interviews with scale developers, educators, and students were conducted to gauge clarity, propriety, and reflection of the constructs of the items.

Construction of the Multiple-Choice Grammar Test: SRSIP. *The Syntactical Relationships as Signaled by Internal Punctuation: Multiple-Choice Grammar Test*

(SRSIP) is a multiple choice test designed to assess respondents' use of internal punctuation and identification of major syntactical units. Scales measure punctuation use of commas, semicolons, and colons and identification of major syntactical structures within the following frameworks.

Grammar Assessment Content. Identification of major syntactical structures (i.e., phrases, independent clauses, and dependent clauses) may provide correlates to punctuation choices made in previous scales of the instrument.

Grammar Assessment Format. Haladyna, Downing, and Rodriguez (2002; see also Haladyna & Downing, 1989a, 1989b) created a taxonomy of 31 "item-writing rules" for the construction of multiple choice tests geared for specific multiple choice formats (p. 312). These rules cover item content, formatting, style, stem-writing, and choice writing. Pertinent to the measure developed for this study was the conventional multiple choice format comprised of a stem and a list of choices, recommended by all textbooks reviewed by Haladyna et al. Whereas the taxonomy was followed in full in the construction of the objective measure developed for this study, rules of particular interest are discussed below.

Types of items in the multiple-choice test. This measure was divided into four sections wherein given sentences were provided with specific directions for task completion and a constant set of choices throughout each scale. Three assessed

punctuation use to mediate syntactical relationships. The fourth section asked respondents to identify syntactical structures (i.e., phrases and clauses).

Question stems. The stems used in SRSIP were sample sentences for which respondents chose the most appropriate action as limited by the choices. Stem forms were constant within each scale of the test. Most important to the writing of the stem was the wording of the directions, which had to clearly reflect the central idea and guide the reader to perform the task correctly. Stem sentences in SRSIP were comprised of common syntactical structures with informal language to avoid unfamiliar vocabulary and trick questions and to better isolate the task in question.

Content. Given the nature of the stems in SRSIP, that they were not written in question form but as sample sentences, content of items focused on the syntactical structure of the sentences rather than on the semantic content. The choices intended to guide the test taker in the mental operations required to perform the task. For example, if the choices were to add a comma, semicolon, colon, or nothing at all at a point in a given sentence, respondents had to parse their understanding of the function of each point of punctuation to mediate the syntax and decide on the best choice.

Distractors. Whereas sentence stems were written with familiar syntactical structure, the points of punctuation or types of syntactical structures listed in the choices presented some possible uses reflecting common deviations from Standard English.

Number of choices. The choices in each scale of SRSIP numbered four and were constant in content per each section of the measure to minimize reading time from item to item. Appendix A presents the structure of SRSIP.

Related survey items. Thirteen items asked respondents to rate their knowledge of syntactical structures (see Table 2 above for items and response choices). Five confidence items asked respondents to rate the degree to which respondents believed their answers mirrored usage of Standard American English (see Table 3 above for items and response choices).

Phase 3: Quantitative Evaluation

Participants. Participants were drawn from a pool of adults with various writing practices professionally and/or academically and personally. No other delimiters existed for participation. A total of 328 people responded to the invitation to participate.

Participants' ages fell between 18 and 77 years ($n = 272$): 24.6% were 18-22 (traditional undergraduate ages); 21.3% were 23-30, 21.3% were 31-40; 18% were 41-50; and 14.7% were 51-77. Ethnicity and race varied ($n = 267$) with the majority being White (86.1%): 2.6% African American, 6.7% Latino, 1.9% Asian, 1.2% American Indian, .9% other.

Out of 273 responding, a majority of participants identified as female (76.9%).

Respondents also reported their professions ($n = 272$): 35.3% in education; 25% in full-time college attendance, 10.3% in business ownership or some other profession; 5.5% in healing arts or counseling; 5.1% in office administration; 4% in library services; 3.3% in writing fields; 3.3% in service industries or skilled trades; 2.9% in the arts; 2.6% in research; 1.5% in the military; and 1.5% in retirement.

Education level of participants was of primary interest. Undergraduate and graduate students of three post-secondary institutions were specifically targeted for participation. Of the 147 respondents who identified as current students, 89.3% reported

attendance at these institutions. The remaining percentage was accounted for through snowball sampling methods. Of those identifying as current students, 46.9% were studying education, 24.8% business, 6.9% library science, 6.9% psychology, 6.2% math or science, 2.1% English studies, and 6.2% other or unreported.

Overall, of the 273 reporting education level, overwhelmingly most had pursued their education beyond high school. No one reported having less than a high school diploma, and only 1.5% reported having completed only high school. The remaining respondents were fairly evenly distributed in their educational levels: 28.6% had some college experience, 24.2% had Bachelor's degrees, 28.2% currently held a Master's degree, and 14.2% either were pursuing or had accomplished degrees or certification beyond Master's level. Among the mix were also the 2.9% who had received a two-year or trade school degree. Given the nature of this sample that would be most likely compelled to participate owing to their interest level in the topic of grammar, it is not surprising that most (68.7%) reported having received grades of A in their high school language arts classes; 28.6% reported averaging Bs with the remainder reporting lower grades. In what kind of secondary institutions did respondents receive those grades? Most (77.7%) reported attendance in a public high school; 9.5% attended a private religious school and 4.8% a private nonreligious school. As for the remaining 8%, they attended alternative schools (1.1%), received their GEDs (1.8%) after attending either public or private institutions, were home schooled (.7%), or experienced a mixture of educational environments (5.1%).

When queried regarding their present writing needs in their professional lives (n = 273), 66.3% reported it as *essential* and another 18.3% counted it as *important*. As for writing in their personal lives (n = 273), their needs varied: 19.8% responded with *essential*; 35.5% with *important*, 30.4% with *somewhat important* and 13.9% as *not important*. In addition, 273 responded to the question regarding the importance of writing in their academic lives: 64.5% deemed it as *essential*, 14.7% as *important*, and the remainder as either *somewhat important*, *not important*, or *not applicable*.

Participants were also asked to rate how much they felt their writing was scrutinized by others (n = 273): 9.5% reported *almost always*; 19.8% reported *more often than not*; 12.8% cited *about half the time*; 23.1% felt scrutinized only sometimes; and most, 34.8%, felt rarely scrutinized. As to whether participants were apt to scrutinize their own writing, responses were in opposition to the responses in the previous question. Very few, 8.1%, rarely scrutinized their own writing, yet 20.1% only sometimes scrutinized it; another 34.7% scrutinized either half the time or more often than not; 36.3%, however, scrutinized their own writing almost always.

Procedure. After receiving IRB approval from the University of Denver, IRB approval was sought from the institutional review board of participating institutions, and a list of institutional email addresses of students was obtained from each. Recruiting emails invited students to pass along the email to other people who might be interested in participating. Additional participants were recruited through snowball sampling methods using social networking sites and other electronic means. The measures were accessed by participants via a survey engine site.

Informed consent was obtained once the participant accessed the website and before the respondent proceeded to complete the measure (see Appendix B). Completion time was 20-25 minutes with the writing sample consuming the greatest amount of time to complete. Confidentiality was assured, and continuing to the next page of the assessment after reading the informed consent page constituted consent. An invitation to provide an email address at completion of the assessments allowed participants to enter into the lottery for a cash reward.

Data collection took four months to complete. Data were collected onto a downloadable dataset and were cleaned and screened in preparation for data analysis. At the end of data collection and analysis, one respondent was randomly selected for award of the incentive, which was sent to the winner on October 7, 2011.

Item Analysis

After data collection was complete, SRSIP items were analyzed for unidimensionality, scale use, item difficulty, discrimination, and construct coverage using the Rasch model. Because items were generated to be multifaceted, multiple dimensions were expected to emerge from the measure as a whole. It was also anticipated that scale items would be correlated and if correlated strongly, a unidimensional structure could be the most parsimonious interpretation of the structure.

Unidimensionality. SRSIP was divided into four test sections: (a) three differently formatted items assessing internal punctuation use (i.e., use of commas, semicolons, and colons) and (b) a test of identification of major syntactical structures (i.e., phrases, independent clauses, and dependent clauses).

The items for internal punctuation utilized varying stems and answer choices. The following grammatical principles were tested in each part of the section on punctuation use:

Rule A: A run on sentence is comprised of two or more independent clauses adjacent to one another without proper punctuation; separate with a semicolon

Rule B: A comma splice is comprised of two or more independent clauses adjacent to one another separated by a comma; use a semicolon instead.

Rule C: No internal punctuation is required for a simple sentence or clause with a compound verb wherein no clause or phrase interactions requiring punctuation exist.

Rule D: When two independent clauses are joined by coordinating conjunction, separate with comma.

Rule E: When an introductory dependent clause precedes independent clause, separate with comma.

Rule F: When an independent clause is followed by a dependent clause introduced with a subordinating conjunction, no comma is needed to separate the two.

Rule G: Use a comma to set off an introductory word or phrase.

Rule H: Use commas to set off appositives and parentheticals.

Rule I: Use commas set off a nonrestrictive relative clause.

Rule J: No commas are needed to set off a restrictive relative clause.

Rule K: No colon is needed if the structures preceding a list cannot stand alone OR a colon is needed if structures introducing items in a list can stand alone.

In the first three sections, the single dimensional structure was expected to emerge based on the punctuation required (i.e., commas, semicolons, or colons) more than on the type of syntactical structure presented (i.e., phrase or clause) because response categories asked respondents to consider types of punctuation use.

The fourth SRSIP test section scale asked respondents to identify three syntactical structures: phrase, independent clause, and dependent clause, and a complex relationship comprised of both independent and dependent clauses and reflect the following grammatical guidelines:

- a. Rule L: A phrase does not have subject-verb relationship.
- b. Rule M: An independent clause contains a subject-verb relationship and stands alone. It may include other clauses as components of the clause pattern, such as a noun clause serving as the direct object.
- c. Rule N: A dependent clause contains a subject-verb relationship but cannot stand alone owing to subordinating conjunctions, relative pronouns, or nominalizers (e.g., *that*) attached to the clause.

A single dimensional structure was expected to emerge reflecting the above rules for phrases, independent clauses, dependent clauses, and a response choice identifying complex syntactical structures (i.e., a combination of independent and dependent clauses). (See Appendix C for test)

Analysis of SRSIP. SRSIP responses were assessed using the Rasch model. SRSIP multiple choice items were scored dichotomously. Items with infit and outfit mean squares that fell out of the range of .5-1.5 (Lincare, 2007) and standardized infit and outfit statistics exceeding |2| (Bond & Fox, 2007) were flagged for further examination as items with fit statistics beyond these ranges were considered less compatible to the model than was expected.

Phase 4: Validation

Upon completion of item analysis, convergent validity was explored by using tests for correlations among related survey items.

Writing Sample. Participants were asked to submit a writing sample wherein they discussed their experiences in learning grammar. They were asked to provide 5 to 7-sentence sample. This sample was assessed specifically for the same types of punctuation use and syntactic structures as assessed in the multiple-choice test. Correlations were calculated between the ratio of total error counts to sentence number in respondents'

writing samples and their performance on the objective test to gauge the content validity of the objective measure. Instructions for the writing sample are as follows:

Please write 5-7 sentences regarding the grammar instruction you have received. Maybe you really enjoyed learning grammar; maybe you did not enjoy it very much but knew it was important to learn; maybe you did not feel that instruction was useful or consistent enough for you to learn it to your satisfaction. In addition, you may remember certain strategies or materials that your instructors used to teach you grammar. Talk about their effectiveness. How does your confidence in your own grammar use impact your writing now, especially in your profession? If you have a story that is pivotal in your experience with grammar use or instruction, please tell it. PLEASE WRITE 5-7 SENTENCES.

Chapter 3

Results

Rasch Analysis of All SRSIP Items

A Rasch analysis of all items was conducted to examine dimensionality and estimate reliability for the test as a whole (see Appendix C).

Dimensionality. Dimensionality of a scale is assessed by examining several indicators. The first are global mean square (MNSQ) and standardized (ZSTD) infit and outfit statistics, which ideally range from 0 to 1 (0.0 expected for standardized) and are suggested by Lincare (2007) to be generally suggestive of unidimensionality if falling into the range of .5 to 1.5 (-3 to +2 for standardized fit). For the full item set, global MNSQ infit was .99 (ZSTD = .1) and outfit was .96 (ZSTD = 0.0). These values fell within suggested ranges.

A second indicator is found in examination of item misfit in addition to correlation (item discrimination) coefficients, which should be positive and substantial. Appendix D presents the misfit order of item difficulty from most misfitting to least as well as correlation coefficients. All fit statistics fell within suggested ranges except for one outfit MNSQ (2.32, ZSTD = 4.8) of item 4-3. Correlation coefficients, however, were positive yet not all substantial, ranging as low as $r = .05$ for item 4-3 and only as

high as $r = .53$ for item 4-6. However, the greater the number of items, the lower the coefficients are expected to be over all.

A third indicator of unidimensionality is an analysis of residuals using Rasch principal components analysis. To begin this evaluation, the variance explained by the measure should be examined. Bond and Fox (2001) suggest that this percentage should not fall below 60%; if it does, then there is evidence of a potential second dimension or component or of misfitting items. In addition, the analysis of residuals can suggest whether a second dimension is indicated once the first dimension (as indicated by total variance unexplained in terms of residuals) has been accounted for. First contrast eigenvalues that exceed 2.0 (Bond & Fox, 2001) or 3.0 (Lincare, 2007) would indicate that more than one dimension is present and, hence, that a supposition of unidimensionality is not supported. The theory here is that an eigenvalue of 2.0 reflects at least two items, the lowest number to suggest the possibility of a second dimension, and 3.0 is most likely indicative of a second dimension (Chiang et al., 2009; Lincare, 2007). It is preferable that the eigenvalue for the 1st contrast be less than 2.0.

Variance explained for the test as a whole scale was 66.2%, which suggested that this particular scale was accounting sufficiently for the variance within responses. However, the eigenvalue 2.5 of the first contrast did indicate a potential second dimension by this analysis, thereby questioning the unidimensionality of the SRSIP test as a whole.

Because evidence of unidimensionality was ultimately lacking for the full set of SRSIP items, the decision was made to test the reliability and dimensionality of two

subscales: Punctuation Items and Phrase/Clause Identification Items. Prior to this decision, a traditional route of deleting items with larger infit and outfit MNSQs proved fruitless in bringing the eigenvalue of 2.5 any closer to 2.0 after five attempts (see Appendix D for first five items). Theoretically, dividing the test into the two sections was more reasonable than the traditional approach, especially as items in Punctuation Items are not discrete; rather, they are contextualized, so, in effect, deleting an item amounts to deleting a sentence from a short paragraph. In Phrase/Clause Identification Items, the sentences appear at least two times with different passages highlighted for identification; item deletion, if necessary, is reasonable (see Appendix C).

Rasch Analysis of SRSIP Punctuation Items (Test Sections 1-3)

A Rasch analysis of Punctuation Items was conducted to test dimensionality and estimate reliability.

Dimensionality. For Punctuation Items, global MNSQ infit was .98 (ZSTD = .1) and outfit was .91 (ZSTD = 0.0). These values fell within suggested ranges.

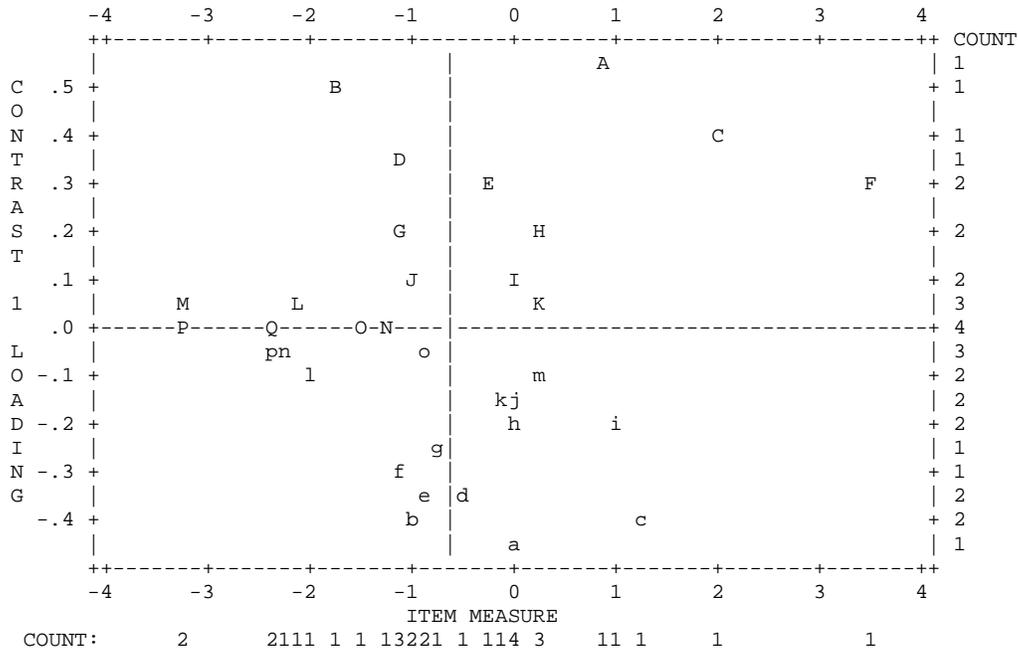
Appendix E presents the misfit order of item difficulty from most misfitting to least as well as correlation coefficients for items in Punctuation Items. All fit statistics fell within suggested ranges except for one outfit MNSQ (1.79, ZSTD = 2.6) of item 1-8. Correlation coefficients between .11 and .50 were positive yet not substantial, owing to the number of items: 34.

Variance explained for this scale was 66.2%, an adequate percentage given the guideline of > 60% (Bond & Fox, 2001). However, the eigenvalue (2.2) of the first contrast gave evidence for a possible second dimension. Even so, Lincare (2007) noted,

“Simulation studies indicate that even Rasch-conforming data produce eigenvalues with values up to 2.0, i.e., with the strength of two items” (p. 247). This eigenvalue of 2.2 suggested that the strength of a possible second dimension was just over the strength of 2 items out of 34. Considering this and that 66.2% of the variance was explained by the measure, it is likely that indicators of a second dimension were a product of noise.

Additional examination of the pattern of residuals for the first contrast did not indicate a second dimension (see Figure 1). If multidimensionality were at play, then as the residuals aligned along the dimensions, first contrast eigenvalues would be higher than those expected from random data. In addition, differences in positively and negatively loading items would be substantive (Lincare, 2007). Figure 1 does not show these substantive differences among items loading positively and negatively; in fact, all loaded positively. Therefore, no evidence of a systematic second dimension was present here.

Figure 1. Standardized Residual Plot for Contrast 1 of Punctuation Items



Moreover, item local independence is indicated by positive yet nonsubstantial coefficients of correlations of the residuals. If coefficients are high, it could mean that the items are repetitive or measuring a second dimension. If they are negative, too much independence is indicated and the items are misfitting. Table 5 presents the largest standardized residual correlations. None are substantive nor are any negative. Again, the evidence here does not suggest a second dimension.

Table 5

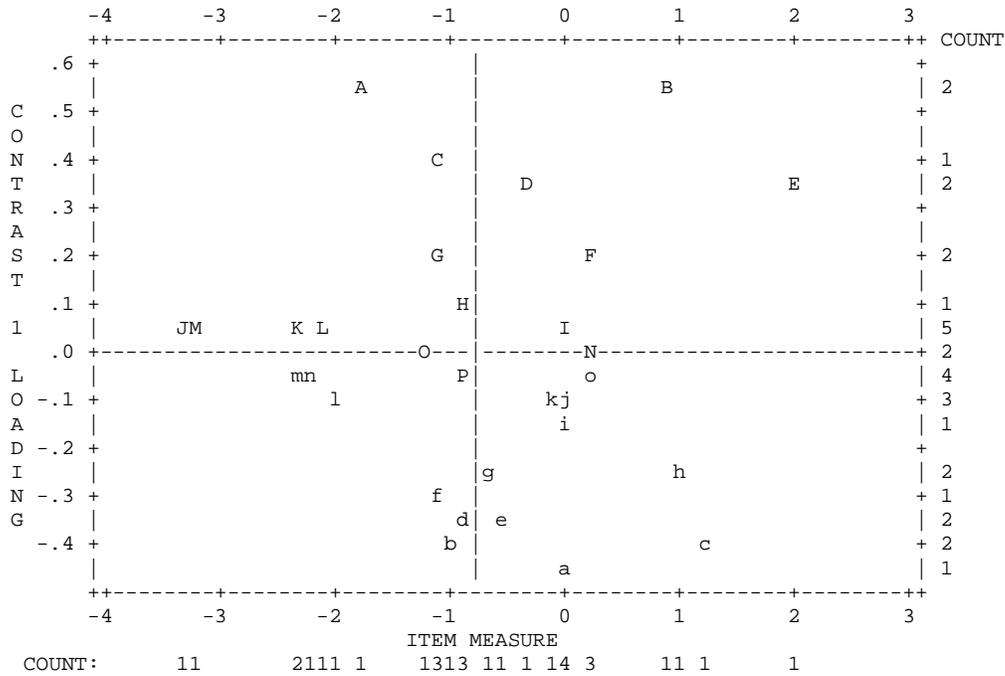
Largest Standardized Residual Correlations for Punctuation Items

Residual Correlation	Item	Item
.36	2-5	3-4
.34	1-6	1-11
.33	1-4	2-10
.29	1-2	2-3
.26	2-4	2-7
.25	1-11	3-3
.25	1-1	2-6
.24	1-11	3-2
.24	1-6	3-2
.21	3-3	3-6

Nevertheless, a traditional approach was taken in a quest to ferret out the possible second dimension by deleting two items with the highest outfit mean squares (all infit mean squares were well within suggested ranges): Item 1-8 (MNSQ 1.79, ZSTD = 2.6) and item 2-2 (MNSQ 1.46, ZSTD = 2.0). Variance explained by the measure remained unchanged. The eigenvalue of the first contrast dropped to 2.1. Whereas this appears to signify an improvement on the surface, examination of the residuals loading plot shows an item now negatively loading (P) whereas, with all items in tact, each loaded positively (see Figure 2).

Figure 2. Table of Standardized Residual Variance (in Eigenvalue units)

		Empirical		Modeled	
Total variance in observations	=	91.8	100.0%	100.0%	100.0%
Variance explained by measures	=	60.8	66.2%	65.5%	65.5%
Unexplained variance (total)	=	31.0	33.8%	100.0%	34.5%
Unexplnd variance in 1st contrast	=	2.1	2.3%	6.9%	6.9%



Most problematic with this approach was that deleting units from a contextualized bank of items interrupts the semantic integrity of the scale. To adjust for this problem, the scale items would have to undergo revision, thereby essentially creating a new test for which the results above would no longer apply.

Given the varied pieces of evidence regarding dimensionality, it was determined that a strong argument for the presence of a second dimension could not be made. Instead for all practical purposes, unidimensionality of Punctuation Items was assumed.

Reliability. Reliability in Rasch modeling is calculated for persons and items from the spread of both in standard error units. This is known as *separation*. Ideally, persons should be spread along the continuum as should items. Neither should be tightly

clustered anywhere along the respective continua. With sufficient spread of both, means should be roughly in line with each other (i.e., person ability to item difficulty).

Separation statistics for both exceeding 1.0 indicate better spread, and from those statistics can reliability of either persons or items be assessed. If separation is low (i.e., closer to 0.0), then reliability will approach 0.0. Of course, the higher the separation and reliability statistics, the more useful the measure (Chiang, Green, & Cox (2009).

Person separation for the entire sample (extreme and nonextreme respondents) was 6.60 with a person reliability estimate of .98. Cronbach's alpha was .96. Likewise, item separation was also high 5.49 with a reliability estimate of .97. Together, these estimates suggested a reliable measure for both persons and items.

Invariance. Scalar invariance is a testable assumption of Rasch modeling, which means the scale items perform similarly across measurement events and across groups. Reported here are the results of a single pilot of the measure to a sample of adults over 18 with expected varying writing needs in their personal and professional and/or academic lives; however, evidence suggested that writing needs of the sample were probably greater as a whole than those in the general population, owing primarily to their education levels (see Participants under Methods). A differential analysis of groups by educational level was conducted to test scalar invariance of Punctuation Items. Education level was grouped as *Some high school* (n = 0); *High school graduate* (n = 4); *Some college* (n = 78); *2-year or trade school degree* (n = 8); *Bachelor's degree* (n = 66); *Master's degree* (n = 77); *Beyond master's degree* (n = 40).

Results of the differential item analysis for Punctuation Items are presented in Appendix F. Of the 204 pairwise comparisons of six groups and 34 items, 3 comparisons, or 1.5%, showed statistically significant differences at $p < .01$. No consistent pattern of differences was apparent. For example, respondents with some college most often differed from other groups, yet the differences were spread randomly across items. For a true DIF to be in play, groups would differ consistently and logically. Results here beg the question that if those with some college are differing on certain items, then why aren't those with less education also differing likewise? Beside the random nature of statistical differences, the percentage of differences found was too small to conclude that items comprising Punctuation Items were not invariant across these respondents' education levels.

Targeting. Figure 3 presents the person/item map for Punctuation Items. The mean person logit ($M = 1.88$, $SD = 1.12$) was higher than that for the items. Ideally, the means should closely align. Whereas item spread was good along the continuum, the persons clustered more closely at the top, producing the higher mean. The difference here shown graphically indicated that the ability of persons exceeded the difficulty of the items to accurately test the construct of use of internal punctuation. A sample of less skilled respondents may be less likely to score as highly as did this sample of people who are generally better educated than the population at large.

Rasch Analysis of SRSIP Phrase/Clause Identification Items (Section 4)

A Rasch analysis of Phrase/Clause Identification Items was conducted to test dimensionality and estimate reliability.

Dimensionality. For the Phrase/Clause Identification Items, global MNSQ infit was .98 (ZSTD = .1) and outfit was .91 (ZSTD = 0.0), all acceptable values. The misfit order of item difficulty as well as correlation coefficients for items in Phrase/Clause Identification Items are presented in Appendix G. All fit statistics fell within suggested ranges except for one outfit MNSQ (2.32, ZSTD = 4.8) of item 4-3. Correlation coefficients were positive and ranging in size as expected for the number of items (17; $r = .21-.54$) with the exception of the very low coefficient of $r = .05$ for item 4-3. Variance explained for this scale was 66.2%. The eigenvalue 2.1 of the first contrast suggested the presence of a possible second dimension.

Given that an actual second dimension may be present, the decision was made to delete item 4-3 as its absence in the test would not impact the contextualized nature of the items given the format; in other words, the test integrity would stay the same. Results of the analysis showed improvement in global MNSQ infit (1.0, ZSTD = .1) and outfit (.98, ZSTD = .1). All item misfit infit and outfit MNSQs were well within suggested ranges (see Appendix G). In addition, the variance explained by the measure remained the same, but the eigenvalue for the first contrast dropped to 2.0. Because deletion of this item was inconsequential to the semantic context of the test, these improvements provided evidence enough to assume that this scale was unidimensional.

Reliability. With the deletion of item 4-3, person reliability estimates and Cronbach's alpha were .69 and .71, respectively. Item separation was 6.48, and reliability estimate was : .98. Whereas item reliability was very good, person reliability was not discountable *per se* but could be improved. Nevertheless, the variability map of the modified scale (see Figure 4) reveals good spread of both items and persons with means of both in close proximity of each other. Strong evidence is not present in this analysis to suggest that the person reliability estimates are too low to consider this test reliable overall, especially in light of the high item reliability estimates.

Invariance. Results of the differential item analysis for Phrase/Clause Identification Items are presented in Appendix H. Of the 96 pairwise comparisons of six groups per 16 items, 7, or 7.3%, showed statistically significant differences at $p < .01$. With a percentage that high, some evidence of DIF was present. However, upon closer examination, group 4 ($n = 8$) statistically differed with groups 3, 5, 6, and 7 on item 4-17, accounting for 4.2% of the DIF. This finding seems anomalous probably owing to the smallness of the group rather than a problem with item itself. Group 4 did not differ from another group on any of the other items. If this group were removed from the analysis, only three items would show differences ($3/96 = 3.1\%$). In that case, one could conclude that Phrase/Clause Identification Items was invariant across these respondents' education levels.

Targeting. As discussed under Reliability of the Phrase/Clause Identification Items, the person/item map (see Figure 4 above) revealed the mean person logit ($M = .09$, $SD = 1.24$) and the item mean to be closely situated with similar patterns of spread along the continuum. This map suggests that these items were capable of assessing the construct of phrase and clause identification for this sample. Keep in mind the high education levels of these respondents, however. If this test were used with a sample of less skilled respondents, the effect might be quite the opposite of the possibilities suggested with the Punctuation Items: persons in that case could be found to cluster more toward the bottom of the continuum, thereby lowering the person logit mean and disrupting the current spread.

Validity Evidence for SRSIP Measures

Distributions. Person logit scores for Punctuation Items and Phrase/Clause Identification Items as well as applicable survey items were screened for univariate normality prior to running bivariate correlations. Normality was assumed when values for skewness and kurtosis did not exceed the generally accepted value of |1.0| when using the IBM SPSS 19 program.

The following logit scores or survey items had skewness and kurtosis values exceeding the guideline: Phrase/Clause Identification Items logits, recognizing run ons; knowing comma use; and total error rate from the writing sample. With the exception of run on totals, these values were not excessive: skewness did not exceed |1.3|, and kurtosis did not exceed |2.3|. Run on values for skewness (-1.6) and kurtosis (2.2) were not excessively out of range and were driven by a few who scored very low ($M = 4.25$, $SD = 1.01$). Means and standard deviations for the test section logit scores, sentence structure and confidence survey items, and writing sample error rate are presented in Appendix I.

Bivariate correlations among validating survey items. In order for survey items and writing sample error rate to be useful in validating the two SRSIP scales, they should intercorrelate as is appropriate with positive, statistically significant, and substantial coefficients (except for correlations with the writing sample error rate for which negative coefficients are expected). Appendix J presents the correlations among various items and/or item sets. With the exception of correlates with the writing sample error rate, all correlations were positive, statistically significant at $p < .001$, and moderate ($r = \sim .3-.6$) to substantial ($r > .6$) with coefficients ranging from $r = .37$ to $.90$.

The error rate was correlated with two item sets: Confidence Item Total ($r = -.11$, not significant) and Sentence Structure Item Total ($r = -.15$, $p < .05$). Such low correlations were not expected but could be owing to the calculation of the error rate itself. The error rate is a ratio of the sum of errors to the sentence number. In and of themselves, each variable had excessively high skewness and kurtosis owing to the range of error counts or number of sentences. Respondents were asked to write 5-7 sentences and most complied ($M = 5.79$, $SD = 1.77$, Skewness = 3.0, Kurtosis = 22.07, Range: 2-22), yet 15% wrote 4 or fewer, and 9% wrote 8 or more with one respondent supplying 22 sentences. The sum of errors ($M = 1.44$, $SD = 1.43$, Skewness = 1.78, Kurtosis = 5.27, Range: 0-10) was also problematic in meeting the assumption of normality primarily because 25% of the sample had no errors, 39% had one, and another 17% had two. One respondent made as many as 10 errors. Correlation among these variables was low ($r = .22$, $p < .001$) because the possibility of these particularly able writers committing more errors did not rise as they continued to add to their writing sample. For example, the error count of the participant who wrote 22 sentences was 3. Conversely, the sentence number of the respondent with 10 errors was 9. Whereas the latter example suggests a possible high correlation between the two, the former example accounts for the weakness of the actual correlation, which was born out by 82% of the sample having 2 or fewer errors. In other words, it essentially did not matter how many sentences respondents wrote; most habitually made few errors over all while a few habitually made many errors. Nevertheless, some accounting for the writing sample was deemed necessary to provide validity evidence to the measure, thus, the creation of the ratio variable. What is not

governable when requesting an authentic writing sample is the preference of the writer to use language structures in the same way that the SRSIP presents language. Sentence variety is an example. Every writer stylistically makes choices to use simple, compound, complex, or compound/complex sentences structures. These choices then give rise to appropriate punctuation use. Better writers, such as those in this sample, tend to vary their structures from sentence to sentence and thereby enrich meaning by careful attention to form; poorer writers are less effective in negotiating meaning within syntactic discourse. Given this constraint, the high ability of the responding writers, and the consequential limitations of the error rate, the use of writing sample to validate SRSIP in this study was minimal.

Correlation between person logit scores for Punctuation Items and Phrase/Clause Identification Items. The correlation coefficient between the two scales was $r = .183$ ($p = .002$, $n = 273$). Overall, person logit scores were higher for Punctuation Items ($M = 1.88$, $SD = 1.12$) than for Phrase/Clause Identification Items ($M = .09$, $SD = 1.24$).

Correlations among person logit scores for Punctuation Items and survey items plus error rate. Person logit scores for Punctuation Items and various configurations of validating survey items were correlated to obtain validity evidence for the measure constructed from Punctuation Items. All coefficients were statistically significant at $p < .001$ and all but two were moderate (see Table 6).

Table 6
Correlations among Punctuation Items Person Logit Scores and Validating Survey Items

Scale or Item	<i>r</i>
Confidence Item Total	.48**
Confidence Punctuation Item Total	.47**
Confidence Sentence Structure Total	.33**
Confidence Comma Use	.34**
Confidence Semicolon Use	.45**
Confidence Colon Use	.43**
Confidence Phrase Use	.31**
Confidence Clause Use	.33**
Sentence Structure Item Total	.47**
Sentence Structure Punctuation Total	.48**
Sentence Structure Components Total	.35**
Recognize Independent Clause	.37**
Recognize Dependent Clause	.35**
Recognize Phrase	.24**
Recognize Run on Sentence	.36**
Recognize Comma Splice	.24**
Know Punctuation between Two Independent Clauses	.43**
Know Punctuation between Independent and Dependent Clauses	.45**
Know when to Use Commas	.31**
Know when to Use Semicolons	.45**
Know when to Use Colons	.43**
Writing Sample Error Rate	.30**

Note. ** $p < .001$

Correlations among person logit scores for Phrase/Clause Identification

Items (minus item 4-3) and survey items plus error rate. Person logit scores for Phrase/Clause Identification Items and various configurations of validating survey items were correlated to obtain validity evidence for Phrase/Clause Identification Items. Coefficients were positive but not all were statistically significant at $p < .001$; all were low (see Table 7).

Table 7
Correlations among Phrase/Clause Identification Items Person Logit Scores and Validating Survey Items

Scale or Item	<i>r</i>
Confidence Item Total	.25**
Confidence Punctuation Item Total	.22**
Confidence Sentence Structure Total	.21**
Confidence Comma Use	.19**
Confidence Semicolon Use	.22**
Confidence Colon Use	.18**
Confidence Phrase Use	.20**
Confidence Clause Use	.21**
Sentence Structure Item Total	.21**
Sentence Structure Punctuation Total	.20**
Sentence Structure Components Total	.17**
Recognize Independent Clause	.16*
Recognize Dependent Clause	.14*
Recognize Phrase	.17**
Recognize Run on Sentence	.11
Recognize Comma Splice	.13*
Know Punctuation between Two Independent Clauses	.17**
Know Punctuation between Independent and Dependent Clauses	.14**
Know when to Use Commas	.19**
Know when to Use Semicolons	.20**
Know when to Use Colons	.20**
Writing Sample Error Rate	-.08

Note. * $p < .05$; ** $p < .001$

One would expect coefficients pertaining to punctuation use (e.g., confidence in or knowledge of comma use) to be higher for Punctuation Items and those pertaining to confidence in and recognition and knowledge of sentence structures to be higher for Phrase/Clause Identification Items. Results do not support that supposition. Whereas coefficients show correlations to be low to moderate among Section 1-3 person logit scores and validating survey items, all correlations were low among those for Phrase/Clause Identification Items logit scores. Respondents overall scored lower on Phrase/Clause Identification Items than they did on Punctuation Items, rating themselves much higher on their confidence as well as recognition and knowledge of sentence punctuation and structures than their performance on the test would indicate. In terms of

their validating function, these survey items fail to capture a functional distinction between performance and self-knowledge of the constructs being assessed.

Chapter 4

Discussion

Governing Theory for Construction of *Syntactical Relationships as Signaled by Internal Punctuation: Multiple-Choice Grammar Test (SRSIP)*

The purpose of this study was to create a reliable and valid objective measure to assess, as the name implies, syntactical relationships as signaled by internal punctuation, specifically commas, semicolons, and colons. Syntactic discourse is regulated by internal punctuation, the monitoring of which depends on the writer's understanding of syntactic structures (i.e., phrases and clauses) in order to most effectively signal semantic relationships clearly and effectively. The points of punctuation assessed in SRSIP were chosen for their less stylistic natures, compared to the dash or the ellipses, and the evidence presented by a wide array of grammarians in the literature, including authors of grammar texts (see Corbett & Finkle, 1992; Kolln, 1984, 2006; Kolln & Funk, 2006, Quirk et al., 1985; Warriner, 1988), indicates that instructional materials present the use of these points of punctuation fairly consistently, depending on the authors' adherence to the "rules" of Standard American English (see Micciche, 2004).

SRSIP is a measure designed to assess objectively rather than through evaluation of authentic writing samples for the purpose of testing conveniently yet effectively (see

Hambleton & Murphy, 1992). Strong arguments, especially by Weaver (1996a, 1996b; Weaver et al., 2001), have been made to support use of students' own writing as the text of grammar because such writing is contextualized and authentic (or as authentic as it can be in an educational environment) (See Alderton & Hughes, 1981). Many critics of objective grammar tests point to the decontextualized nature of the language (see Alderton & Hughes, 1981). Peruse most grammar texts (e.g., Warriner, 1988), and one will find tests filled with sample sentences for which the student must make some kind of decision as to the item's correctness, either right or wrong. But the items traditionally are discrete, that is, not contextualized, and thereby are less translatable into the student's own developing sense of language (see Weaver above). The theory guiding the development of SRSIP was to objectively test using a series of contextualized sentences, such as anecdotes, to bridge the gap in the literature between traditional grammar test items and use of student writing samples to assess grammatical competency.

The texts for the SRSIP test sections were written in conversational rather than formal language using three stories and an advertisement (see David, 2007). Each sentence test item built semantically on the previous sentences. Conversational language can be as complex in structure as is academic discourse. The language of SRSIP test texts were intended to appeal to a broad audience who might not share the experiences of the narrator or be interested in answering the advertisement yet would find the content unthreatening, which could lessen the distraction that more formal language might impose on the testing event (see Appendix C).

Necessarily, for the sake of clarity in measuring discrete grammatical constructs, the decision was made to develop a measure testing a narrow yet arguably important set of grammatical concepts as delineated generally by grammarians and anecdotally by learners: punctuation use within the sentence, specifically, commas, semicolons, and colons. Peruse grammar books further, and it is common to see that internal punctuation traditionally has not been taught emphatically in direct conjunction with syntactic structures, hence the SRSIP test section in recognizing these structures. The reality is that uses of commas, semicolons, and colons are inextricably linked to one's understanding of sentences structures (i.e., phrases and clauses). The uses for any of those points of punctuation are much more nuanced than those of, say, apostrophes, a narrow construct in terms of application, yet most traditional grammarians conceptually lump all points of punctuation together and thus assess the constructs.

Evaluation of SRSIP Test Texts

The test texts for the three sections assessing punctuation were alike in form in that each sentence item semantically followed the previous ones. Section 3, however, veered from the anecdote as text form; the text here was in the form of an advertisement, and the language and content were somewhat more formal than that of the others.

In the fourth test section, a different approach was taken. The entire test text was presented first in its entirety as were definitions for the assessed items: phrase, independent clause, dependent clause, and a complex interaction of independent and dependent clauses. Unveiling of items did not proceed as before with each item discretely presenting a semantic follow to the preceding sentences. Instead, sentences often

reappeared once or twice before proceeding semantically. In those repeated items, different structures were highlighted for the respondent to evaluate (see Appendix C). Performance on this section was considerably lower than it was on the other three sections. Two possible explanations are salient: (a) despite the provision of definitions, perhaps respondents were unable to confidently recall the meaning of *subject-verb relationship*, thereby hindering their ability to accurately recognize the assessed structures, and (b) the formatting of the items may have been distracting because it detoured too significantly from that of the items in the previous sections. Revision of the text for Phrase/Clause Identification Items to conform to the formatting of the previous sections is a possibility; however, even though results of the Rasch analysis show reliability estimates that are lower than those for Punctuation Items, revision is not necessarily warranted.

Theory behind SRSIP Response Choice Sets

Response choices are another crucial consideration of objective test construction. According to some psychometricians (see Haladyna & Downing, 1989 a, 1989b; Haladyna et al., 2002; Rodriguez, 2005), the most psychometrically effective response choices include one right answer, partially correct answer(s) as distractors, and wrong answers. All answers should fall within the realm of plausibility, however. For example, all SRSIP response choice sets offered choices regarding either punctuation with commas, semicolons, or colons or identification of phrases and independent and dependent clauses in keeping with the instructions for each test section. The intention of repeating response choice sets per test section was to maintain the respondents' focus

throughout the testing process to lessen measurement error. In the first section, respondents were asked to evaluate whether punctuation was missing from the sentence. In the second section, respondents were asked to evaluate whether punctuation should be added to clarify semantics. In the third section, testers were asked to focus on comma use, whether **all** commas should be retained, removed, or replaced by semicolons. (Colon use was not assessed in this section as colon and comma uses are not often confused.). In the fourth section, respondents simply had to choose among types of syntactic structures to identify highlighted structures.

Dimensionality of SRSIP Scales: Punctuation Items and Phrase/Clause

Identification Items

There are four primary indicators of unidimensionality provided in the Winsteps software: (a) global MNSQ (and standardized fit) infit and outfit statistics: MNSQ = .5-1.5 (ZSTD = -3 to +2); (b) item misfit statistics, whose ranges mirror global infit/outfit ranges, and associated correlation coefficients, which should be positive and substantial; (c) variance explained by the measure (> 60%); and (d) eigenvalues of the first contrast (< 2.0) (see Bond and Fox, 2001; Lincare 2007). Whereas the eigenvalue of the first contrast for all items of the Phrase/Clause Identification Items was 2.2, deletion of one item brought down the eigenvalue to 2.0, and the scale was deemed unidimensional. Dimensionality of Punctuation Items was harder to decipher. The eigenvalue of the first contrast, 2.2, was resistant to change and largely inexplicable in its nature upon examination of all the evidence. Item deletion proved ineffectual as well as irrational. Each section was analyzed independently, and each analysis produced estimates that

generally fell within suggested guidelines. Why the combination of the sections gave evidence of a second dimension is up for debate. And one might ask why advocate that all three sections be considered a single scale when evidence of unidimensionality easily exists for the independent sections. The answer lies in the types of punctuation use appearing in each section. Different uses of commas, semicolons, and colons are assessed across the three sections but may cluster somewhat in individual sections or not appear at all, such as the absence of colon assessment in Section 3. The test is a more well-rounded assessment with all three sections together. When deciding whether to consider Punctuation Items unidimensional or not lay in examination of all the evidence, and the evidence did not support a second dimension.

Reliability of SRSIP Scales and Sample Consideration

Person separation and reliability estimates were very strong for SRSIP Punctuation Items, mirroring item estimates. Person separation and reliability estimates were not as strong for Phrase/Clause Identification Items (.69, Cronbach's alpha = .71), yet other indicators (see Figure 4) did not give evidence that persons were clustering either at the top or the bottom of the continuum as means for persons and items were proximate to each other. Perhaps a sample comprised of adults with more varied writing needs and educational differences would allow more accurate testing of the instrument. This sample was largely well-educated and interested in the topic.

Validity of SRSIP Scales: Veracity of Writing Sample and Survey Items as Useful

Validation Tools

As discussed earlier, the writing sample held promise of offering validity evidence to the SRSIP scales, but the strategy used to quantify the respondents' language use comparable to that assessed in SRSIP was unsuccessful given the nature and limitations of the variable created from an authentic writing sample for comparison to scale scores.

The survey items constructed for this study were designed so as not to bog down respondents with too many grammatical terms as SRSIP sought to test a narrow set of syntactical constructs and their adjacent punctuation. The detail of the terms was limited to *comma, semicolon, colon, phrase, independent and dependent clauses*. Many respondents, especially when further away in proximity to their direct grammar instruction, may not necessarily have recalled even these few terms but still may have known how to punctuate for clarity's sake. Correlations among the validating survey items and the person logit scores were generally low to moderate; thus, their use here was only somewhat promising.

Limitations

The primary limitation of this study was in the recruitment of participants. Respondents were primarily those with higher levels of education than that of the general population. If SRSIP subscales are designed to assess grammatical skills supposedly learned under a wide array of educational conditions, then a sample of learners with more

varied skills could provide greater insight into the effectiveness of the measures created for this study.

An additional limitation is in the construction of items themselves. What constitutes a good item when the item is so dependent on subtleties of language rather than on content? Put that item into a contextual story, and the possibilities of choice compound. This has always been the bane of any kind of grammatical testing. No matter what the language, the meaning of what is being conveyed takes precedence in the mind of the reader. So then how can a test be constructed that is meaningful when it seems most reasonable to present what is most banal to dissipate strong associations within language in order to cut through to the actual constructs being tested, such as internal punctuation or identification of phrases and clauses? The task begs all the more the arguments against objective testing because inauthentic writing depletes the cognition of meaning on the part of the reader/test taker. This depletion will occur no matter how strongly the arguments are made to test. Language and meaning likely override all other associations during a grammatical testing event regardless of how great the implications of performing well are presented to the test taker. This is all to say that perhaps better stories could have been composed, but in order to be truly effective, the stories would have to be tailored for specific samples, which presents inefficient conditions on test analysis in terms of providing evidence of unidimensionality and reliability, especially if the test is to be adopted to be used on larger scales.

Further Research

Exploring invariance. The purpose of this study was to explore the creation of a reliable and valid assessment of a narrow set of grammatical skills by using a unique format (discussed previously) that could be suitable for testing other grammar skills, such as other forms of punctuation, parallel construction, tense agreement, etc. Rasch analyses of the two SRSIP scales showed promising reliability and unidimensionality for each of the scales. Results of the DIF analysis gave minimal evidence that either scale was differentiating among groups on the most appropriate variable assessed: educational level. Should SRSIP be readministered to a sample of students more varied in skill level and age as discussed previously and then reanalyzed, then scalar invariance could be evaluated further to obtain such information as to which groups the language and content of the test most appeal given age and skill level.

Results for unidimensionality and reliability of both persons and items as well as scalar invariance were found to be promising in these SRSIP scales. It is possible then to consider creating additional tests using the SRSIP formatting. The purpose of the test is the primary consideration at this point. Under what circumstances could SRSIP or a SRSIP-like bank of tests be used most effectively? As either formative tests or summative tests in a set of instruction materials? Could such a bank of tests be appropriate for large scale assessment, such as standardized state tests? Determining the purpose of the test is most reliant on deciding which students are the intended audiences. Here, further analysis of scalar invariance would prove most useful.

Tailoring item construction for specific samples. In its current form, SRSIP was tailored for the sample that responded to the invitation to participate: undergraduate and graduate students in three universities in a mountainous state (47% of participants) as well as adults, most of whom either presently reside in the state or have in the past, hence, the stories involving snow and skiing and the advertisement for a college journalism post. These stories narrow the relevancy for potential test takers from other regions of the country, for example, where snowfall or skiing is not common, or for nonstudents. The story involving the drive to the airport may have broader appeal in terms of experience, but its content reflects circumstances more likely to be shared by adults rather than younger people.

The question then arises as to whether this test would be appropriate for other samples. Probably not if the goal is to present contextualized language, assuming that the contextualization of the language would also include attention to content in addition to form. The strength of this study was to test a unique assessment format, the results of which showed promising psychometric properties. Construction of the test is largely formulaic, however. One must choose the grammatical constructs to be tested and decide on the overall length of the test to guide the length of each story. For example, in the case of colon and semicolon use, only a very few uses of those points of punctuation are called for within the guidelines of SAE; therefore, fewer items were needed to test their use in SRSIP. Comma use, on the other hand, is more varied; its use clarifies many relationships among phrases and clauses, thus requiring more items. To keep the length of Punctuation Items scale to a minimum, each assessment of a particular punctuation use consisted of

two or three items. The same strategy was used in constructing the Phrase/Clause Identification Items. Once the decision was made as to what points of punctuation use or syntactic structures to assess, then item construction proceeded to match response format. In terms of content of the items, tailoring the stories to appeal to a broader audience or younger audience, etc., is not problematic because of the construction formula. Further exploration of item construction is warranted.

Continued quantitative analysis of the writing sample. The writing sample yielded rich and ample text to analyze further. Sample characteristics may be quantitatively analyzed with the current writing sample data, expanding to include types of sentences (simple, compound, complex, and compound/complex).

Teaching methods. The content of the writing samples revealed respondents' past experiences in grammar instruction. Qualitative study of these texts especially in conjunction with sample characteristics is warranted, especially in consideration of this group who voluntarily agreed to participation most likely because of higher interest in the topic than is found in the general population. Overwhelmingly, the respondents reported very positive learning experiences, which for the audience of writing and language arts teachers could yield valuable information as to effective methods of teaching grammar.

Validation of SRSIP. Further exploration of means to validate an objective measure assessing grammar, specifically SRSIP, is needed beyond correlating with other grammar measures, most of which are not validated themselves or the results of their IRT analyses and validation processes are not available to the public, (e.g., state standardized tests) (see Michael & Schaffer, 1979; Wyse, 2001). However, if a test could be found that

has undergone IRT analysis and validation processes, it could offer additional valuable insight into the validity of SRSIP beyond what the validating survey items in this study could provide.

Conclusions and Implications for Practice

Assessing objectively to evaluate such nuanced understanding of writers' negotiation of syntactical relationships by punctuation use is a daunting task because the process of deciding what to test and how to test are so fraught with a maddening array of continually asked questions. What are the rules governing such usage and whose rules apply? Does the writer's adherence to some kind of standard for communicating ensure clear understanding on the part of the reader? Are the rules of SAE invariant over time and across cultures (e.g., what about antiquated SAE rules regarding the ban on splitting infinitives, a notion based on Latin grammar—at what point do the rules become antiquated)? Of what use is that standard if both reader and writer entertain different notions as to what is being communicated by, say, use of internal punctuation? The ultimate question is this: is communication clearly conveyed without the use of a grammatical standard? Then the question arises: how do we ensure that purveyors of the standard (i.e., teachers) are familiar with the standard and in agreement about its usefulness? If teachers do not agree, what can we expect when assessing objectively or otherwise according to the standard? So many questions, hence, the debate.

Somewhere lost in the debate are the actual skills that people of all writing needs use to communicate. Those skills grow more eclectic the further away from a standard the writer veers. The people who volunteered to participate in this study for the most part

prided themselves not only in knowing the standard but adhering to it as well. In their writing samples, many reported either strong parental influence in their learning the standard—many of those parents were teachers—or they reminisced about those teachers who directly taught grammar. Those who recollected inconsistent or even nonexistent instruction for most part scored lowest on SRSIP.

If we as a culture agree that a standard for grammar should be taught consistently and that that standard should be SAE, then there is a use for tests like SRSIP. Otherwise, without an endorsement of such an agreement by academic voices in the teaching of writing (e.g., CCCC or NCTE) (see CCCC, 1974; NCTE, 1985, 1994), the status quo of inconsistent teaching of a standard will endure at all educational levels, including those programs wherein students are trained to teach language arts. Consequently, the creation of objective grammar tests will prove fruitless because their use will not be embraced by educational institutions or by students themselves.

So what do we make of this exploration in the creation of an objective grammar test? Using the most sophisticated analyses available to date, the SRSIP scales demonstrated properties of unidimensionality, reliability, and invariance. The purpose of the study was achieved: that objective measures could be created that adequately assess narrow grammatical constructs, which is a necessary condition for such tests, especially one formatted as is SRSIP. These results of this study suggest that constructing a bank of tests with reliable and unidimensional scales using the same format as SRSIP's is achievable. Whereas carefully constructing objective tests is labor intensive, their use on large scale assessments is cost efficient (Cooper, 1980). Cooper also suggested that using

both objective tests and evaluation of authentic writing samples was optimal but not entirely efficient. Halpin et al's (1981) study supported this viewpoint in their findings that the variance accounted for by the objective test alone was relatively low (29%). What was not addressed in the study were the psychometric properties of the Missouri English Test as analyzed using Rasch analysis, so it is unclear as to how good the test really was. Using the Test of Standard Written English (TSWE) as their measure of grammatical competency, Michael and Shaffer (1979) came to the opposite conclusion as Halpin et al. The TSWE was equally predictive of academic success as were the scored essays. Again, however, the psychometric properties of the TSWE were not discussed fully, so it is difficult to conclude that using only objective measures is warranted, that is, until we know how good the test is. In this study, two psychometrically sound scales were created using contextualized language to bridge the gap between objective and authentic language. Given these test properties, a new argument against their use is not readily conceivable.

Implications. The skills assessed in the SRSIP scales are those most problematic not only for teachers to instruct but for students to understand as well. Without a clear understanding of syntactic structures (i.e., phrases and clauses) and how meaning presented in these structures is negotiated through the use of internal punctuation (e.g., commas, semicolons, and colons), clarity is unattainable for the writer and, hence, the reader. Even though the two SRSIP scales parse out the use of internal punctuation and identification of syntactic structures, the theory driving the SRSIP scales is that these constructs are inextricable. For a teacher to use the scales to assess student understanding

of these skills is to presume that the two constructs have been taught in tandem. For example, how is a student to understand that for the sake of clarity it is appropriate to place a comma between two independent clauses joined by a coordinating conjunction if that student has no foundational understanding of syntactic structures? With regard to instructional materials and strategies, this construct should be the foundation upon which all grammatical instruction is built. Grammar instruction must necessarily start with the subject-verb relationship, the presence of it in clauses and the absence of it in phrases. Then instruction can move on to conjunctions and their power to coordinate or subordinate. Next add punctuation. And so on. With that knowledge of syntactic structures in place, a learner is better equipped to see how all other grammatical concepts fall within the hierarchy of the language's construction and functioning. The SRSIP scales are designed to reflect that hierarchy, but teacher adherence to this theory of grammar instruction is necessary for the test to be of real use. Consequently, use of this test as an outcome measure would necessarily have to be accompanied by a set of instruction materials tailored to the theory or presentation of theory through the use of existing materials, which is altogether possible.

Given the level of language used in the SRSIP scales, this test is appropriate for adults or post-secondary students, but as discussed earlier, the content of the stories is perhaps too limited to be relevant to a broader population. Perhaps in a revised form, this test could be incorporated into a large scale test, such as a state-mandated test, or used in college admissions procedures. Its use then would perhaps compel secondary educators to teach the constructs more carefully and methodically than is seen now. However, creation

of a series of instructional materials presenting this theory of grammar instruction to be used from elementary through secondary school would be ideal. Practically speaking, a test is useful in one administration only. Tests must be constructed continually. What this study has shown is that good objective grammar tests can be constructed, that if the instruction has prepared the student for the test, then the test results can be trusted to be reliable within the parameters of our current notions of test reliability.

The primary revelation in creating such measures is that the debate in the literature regarding objective testing of grammar skills can move forward from its current stasis. Perhaps this contribution could then help shape the debate regarding the notion of a standard grammar as a respected anchor for communication throughout society on a large scale. The idea here is that if a standard grammar is testable—assuming that only an agreed upon standard is testable because a set of rules exists—perhaps then it is actually useful to embrace it. The crux of the debate in the literature is the teaching of that standard, that is, the expectation of students to learn to use standard grammar. That means that teachers have to know it themselves and be trained and equipped with good materials in the instruction of it. At least with efforts like these to make a valid and reliable test, ideas among educators and grammarians can continue to be exchanged; at best, the implications of the ability to assess conveniently with objective tests may inspire these same debaters to further explore instructional methodologies using assessments, like SRSIP, that are research-based.

References

- ACT. (2004). National Curriculum Survey calibrates ACT's assessment tools. *Activity*, 42(1). Retrieved February 1, 2008, from <http://www.act.org/activity/winter2004/survey.html>
- Alderson, J. C. (1981a). Reaction to Morrow paper (3). In J. C. Alderton & A. Hughes (Eds.), *Issues in language testing* (ELT documents 111) (pp. 45-54). London, England: British Council, English Language Division.
- Alderson, J. C. (1981b). Report of the discussion on communicative language testing. In J. C. Alderton & A. Hughes (Eds.), *Issues in language testing* (ELT documents 111) (pp. 55-65). London, England: British Council, English Language Division.
- Alderson, J. C., & Hughes, A. (Eds.). (1981). *Issues in language testing* (ELT documents 111). London, England: British Council, English Language Division.
- Andrews, R., Torgerson, C., Beverton, S., Freeman, A., Locke, T., Low, G., et al. (2006). The effect of grammar teaching on writing development. *British Educational Research Journal*, 32(1), 39-55.
- Atwell, N. (1987). *In the middle: Writing, reading, and learning with adolescents*. Portsmouth, NH: Heinemann.
- Basset, P. F. (1981, April). English grammar—Can we afford not to teach it? *NASSP Bulletin*, 55-63.
- Bateman, D. R., & Zidonis, F. J. (1966). *The effect of a study of transformational grammar on the writing of ninth and tenth graders*. Urbana, IL: National Council of Teachers of English.

- Beil, C., & Knight, M. A. (2007). Understanding the gap between high school and college writing. *Assessment Update: Progress, Trends, and Practices in Higher Education*, 19(6), 6-8.
- Benjamin, A. (2006). Teacher to teacher: What is your most compelling reason for teaching grammar? *English Journal*, 95(5), 18-21.
- Bond, T. G. & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Braddock, R., Lloyd-Jones, R., & Schoer, L. (1963). *Research in written composition*. Urbana, IL: National Council of Teachers of English.
- Brown, A. R. (1996). Correct grammar so essential to effective writing can be taught--really! *English Journal*, 85(7), 98-101.
- Chiang, K. S., Green, K. E., & Cox, E. O. (2009). Rasch analysis of the Geriatric Depression Scale – Short Form. *The Gerontologist*, 49(2), 262-275.
doi:10.1093/geront/
- Chidomere, R. C. (1989). Test item arrangement and student performance in Principles of Marketing examination: A replication study. *Journal of Marketing Education*, 40, 36-40.
- Chomsky, N. (1957). *Syntactic structures*. The Hague/Paris: Mouton.
- Colorado Department of Education. (1995). *Colorado model content standards: Reading and writing*. Retrieved November, 18, 2007, from <http://www.cde.state.co.us/cdeassess/documents/OSA/standards/reading.pdf>

- Combs, W. E. (1976). Further effects of sentence-combining practice on writing ability, *Research in the Teaching of English*, 10, 137–149.
- Combs, W. E. (1977). Sentence-combining practice: Do gains in judgments of writing “quality” persist? *Journal of Educational Research*, 70(6), 318–321.
- Conference on College Composition and Communication. (1974). Students' right to their own language. *College Composition and Communication*, 25, 1-25.
- Cooper, P. L. (1984). *Assessment of writing ability: A review of research*. GRE Board Research Report (No. 82-15R). Princeton, NJ: Educational Testing Services.
- Corbett, E. P. J., & Finkle, S. L. (1992). *The little English handbook* (6th ed.). New York: Harper Collins.
- Crowhurst, M. (1983). Syntactic complexity and writing quality: A review. *Canadian Journal of Education*, 8(1), 1-16.
- Daiker, D. A., Kerek, A., & Morenberg, M. (1978). Sentence-combining and syntactic maturity in freshman English. *College Composition and Communication*, 29(1), 36-41.
- Dávid, G. (2007). Investigating the performance of alternative types of grammar items. *Language Testing*, 24(1), 65–97.
- Dawkins, J. (1995). Teaching punctuation as a rhetorical tool. *College Composition and Communication*, 46(4), 533-548.
- Devet, B. (2002). Welcoming grammar back into the writing classroom. *TETYC*, 30(1), 8-17.

- Elbow, P. (1986). *Writing with power: Techniques for mastering the writing process*. New York: Oxford University Press.
- Elley, W. B., Barham, I. H., Lamb, H., & Wyllie, M. (1975). The role of grammar in a secondary school English curriculum. *New Zealand Journal of Educational Studies, 10*, 26–42.
- Faigley, L. (1980). Names in search of a concept: Maturity, fluency, complexity, and growth in written syntax. *College Composition and Communication, 31*(3), 291-300.
- Fearn, L., & Farnan, N. (2007). When is a verb: Using functional grammar to teach writing. *Journal of Basic Writing, 26*(1), 63-87.
- Freire, P., & Shor, I. (1987). *Pedagogy for liberation: Dialogues on transforming education*. London: Macmillan.
- Green A. B., & Weir, C. J. (2004). Can placement tests inform instructional decisions? *Language Testing, 21*(4), 467–494.
- Haladyna, T. M., & Downing, S. M. (1989a). A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education, 2*(1), 37-50.
- Haladyna, T. M., & Downing, S. M. (1989b). Validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education, 2*(1), 51-78.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309–334.

- Halpin, G., Halpin, G., & Schaer, B. B. (1981, March). Research on writing: A search for objective measures related to holistically scored essays. Paper presented at the Annual Meeting of the Eastern Educational Research Association. Philadelphia, PA.
- Hambleton, R. K., & Murphy, E. (1992). A psychometric perspective on authentic measurement. *Applied Measurement in Education*, 5(1), 1-16.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer.
- Hannay, M., & Kroon, C. (2005). Acts and the relationship between discourse and grammar. *Functions of Language*, 12(1), 87-124.
- Hartwell, P. (1985). Grammar, grammars, and the teaching of grammar. *College English*, 47(2), 105-127.
- Hawkins, C. W. (1987). Use of the Rasch model in communication education: An explanation and example application. *Communication Education*, 36, 107-112.
- Hoffman, M. (2006, Spring). More grammar gaps. *Academic Exchange*, 220-225.
- Hunt, K. W. (1965) *Grammatical structures written at three grade levels*. NCTE Research Report No. 3. Champaign, IL: National Council of Teachers of English.
- Hunt, K. W. (1970). Syntactic maturity in schoolchildren and adults. *Monographs of the Society for Research in Child Development*, 35(1), iii-iv & 1-67.
- Hunt, K. W., & O'Donnell, R. (1970). *An elementary school curriculum to develop better writing skills*. Washington, DC: Office of Education, Bureau of Research.

- Hutchinson, D., McCavitt, S., Rude, K., Vallow, D. (2002). *Improving student achievement through grammar instruction*. Chicago: Master's of Arts Action Research Project, Saint Xavier University.
- Kolln, M. (1981). Closing the books on alchemy. *College Composition and Communication*, 32(2), 139-151.
- Kolln, M. (1983). Reply by Martha Kolln. *College Composition and Communication*, 34(4), 496-500.
- Kolln, M. (1984). *Language and composition: A handbook and rhetoric*. New York: Macmillan.
- Kolln, M. (2006). *Rhetorical grammar: Grammatical choices, rhetorical effects* (5th ed.). New York: Longman.
- Kolln, M., & Funk, R. (2006). *Understanding English grammar* (7th ed.) New York: Longman.
- Kolln, M., & Hancock, C. (2005). The story of English grammar in United States schools. *English Teaching: Practice and Critique*, 4(3), 11-31. Retrieved November 20, 2007, from <http://education.waikato.ac.nz/research/files/etpc/2005v4n3art1.pdf>.
- Kourlourianos, E., & Marienau, V. (2001). Improving students' confidence in writing achievement through the use of parent, student, and teacher interventions. Chicago: Master's of Arts Action Research Project, Saint Xavier University.
- Lincare, J. M. (2007). A user's manual to Winsteps/Ministeps: Rasch-model computer programs. Chicago, IL: Winsteps.

- MacDonald, P., & Paunonen, S. (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement, 62*, 921-943.
- Marzano, R. J. (1976). The sentence combining myth. *The English Journal, 65*(2), 57-59.
- Mellon, J. C. (1969). *Transformational sentence-combining: A method for enhancing the development of syntactic fluency in English composition*. Champaign, IL: National Council of Teachers of English.
- Mellon, J. C. (1981). *Sentence-combining skills, results of the sentence-combining exercises in the 1978-79 national writing assessment: Prepared by the National Assessment of Educational Progress*. Denver, CO: Education Commission of the States.
- Micciche, L. R. (2004). Making a case for rhetorical grammar. *College Composition and Communication, 55*(4), 716-737.
- Michael, W. B., & Shaffer, P. (1979). A comparison of the validity of the Test of Standard Written English (TSWE) and of the California State University English Placement Test (Csuc-Ept) in the prediction of grades in a basic English composition course and of overall freshman-year grade point average. *Educational and Psychological Measurement, 39*, 131-145. doi: 10.1177/001316447903900118
- Moffet, J. (1968). *Teaching the universe of discourse*. Boston: Houghton Mifflin.

- Moller, A. (1981). Reaction to Morrow paper (2). In J. C. Alderton & A. Hughes (Eds.), *Issues in language testing* (ELT documents 111) (pp. 38-44). London, England: British Council, English Language Division.
- Morrow, K. (1981). Revolution or evolution? In J. C. Alderton & A. Hughes (Eds.), *Issues in language testing* (ELT documents 111) (pp. 9-25). London, England: British Council, English Language Division.
- National Council of Teachers of English. (1985). *Resolution on grammar exercises to teach speaking and writing*. Retrieved April 12, 2006, from <http://www.ncte.org/about/over/positions/category/gram/107492.htm>
- National Council of Teachers of English. (1994). *Resolution on language study*. Retrieved April 12, 2006, from <http://www.ncte.org/about/over/positions/category/gram/107490.htm>
- Nunan, S. L. (2005). Forgive ourselves and forging ahead: Teaching grammar in a new millennium. *English Journal*, 94(4), 70-75.
- O'Hare, F. (1973). *Sentence combining: Improving student writing without formal grammar instruction*. Champaign, IL: National Council of Teachers of English.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English Language*. London: Longman.
- Petit, A. (2003). The stylish semicolon: Teaching punctuation as rhetorical choice. *The English Journal*, 92(3), 66-72.
- Petrovitz, W. (1997). The role of context in the presentation of grammar. *ELT Journal*, 51(3), 201-206.

- Rapeer, L. W. (1913). The problem of formal grammar in elementary education. *The Journal of Educational Psychology*, 4(3), 133-137.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*.
Copenhagen: Danish Institute for Educational Research.
- Rodriguez, M. C. (2005, Summer). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 3-13.
- Rose, S. K. (1983). Down from the haymow: One hundred years of sentence-combining. *College English*, 45(5), 483-491.
- Schleppegrell, M. J. (1998). Grammar as resource: Writing a description. *Research in the Teaching of English*, 32(2), 182-211.
- Scherff, L., & Piazza, C. (2005). The more things change, the more they stay the same: A survey of high school students' writing experiences. *Research in the Teaching of English*, 39(3), 271-304.
- Strong, W. (1976). Sentence combining: Back to basics and beyond. *The English Journal*, 65(2), 56 & 60-64.
- Vavra, E. (1996). On not teaching grammar. *English Journal*, 85(7), 32-37.
- Vavra, E. (2003). Grammar is back, but when will we start cooking? *English Journal*, 92(3), 86-89.
- Warriner, J. (1988). *English composition and grammar: Complete course*. New York: Harcourt Brace Jovanovich.

- Weaver, C. (1996a). Teaching grammar in the context of writing. *English Journal*, 82(5), 15-24.
- Weaver, C. (1996b). *Teaching grammar in context*. Portsmouth, NH: Heinemann.
- Weaver, C., McNally, C., & Moerman, S. (2001). To grammar or not to grammar: That is *not* the question! *Voices from the Middle*, 8(3), 17-33.
- Weir, C. J. (1981). Reaction to Morrow paper (1). In J. C. Alderton & A. Hughes (Eds.), *Issues in language testing* (ELT documents 111) (pp. 27-37). London, England: British Council, English Language Division.
- Wiltse, E. M. (2001, August). *The effects of motivation and anxiety on students' use of instructor comments*. Paper presented at the Annual meeting the Association for Education in Journalism and Mass Communication, Washington DC.
- Wyse, D. (2001). Grammar. For writing? A critical review of empirical evidence. *British Journal of Educational Studies*, 49(4), 411-427.
- Wyse, D. (2006). Pupils' word choices and the teaching of grammar. *Cambridge Journal of Education*, 36(1), 31-47.
- Zimmerman, B. J., & Risemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology*, 22, 73-101.

Appendix A

Test Sections for SRSIP

Subscale	Directions	Response Choices	Number of Items
Section 1	For each of the following sentences, choose one answer that best clarifies the sentence's meaning from the choices given.	A. The sentence is missing a comma. B. The sentence is missing a semicolon. C. The sentence is missing a colon. D. No additional punctuation is needed in the sentence.	12
Section 2	Locate the caret (^) in the following sentences and choose the answer that best clarifies the sentence's meaning from the choices given.	A. Insert a comma. B. Insert a semicolon. C. Insert a colon. D. No additional punctuation is needed.	11
Section 3	For each of the following sentences, choose one answer that best clarifies the sentence's meaning from the choices given.	A. Retain all commas. B. Remove all commas. C. Replace comma(s) with semicolon(s).	11
Section 4	<p>In each sentence below, some of the words are underlined. Circle the answer that best approximates your understanding of phrases and clauses.</p> <p><i>Phrase</i> is defined as a unit of words that does not contain a subject-verb relationship.</p> <p><i>Independent Clause</i> is defined as a unit of words that does contain a subject-verb relationship and can stand alone. Clauses remain independent if they are joined by coordinating conjunctions (e.g., <i>and, or, but</i>, etc.) or conjunctive adverbs (e.g., <i>however, furthermore, therefore</i>, etc.).</p> <p><i>Dependent clause</i> is defined as a unit of words that does contain a subject-verb relationship yet cannot stand alone. Clauses most often are made dependent when paired with subordinating conjunctions (e.g., <i>because, while</i>, etc.),</p>	<p>A. The underlined words constitute phrase(s) only.</p> <p>B. The underlined words constitute an independent clause perhaps with modifying phrases.</p> <p>C. The underlined words constitute a dependent clause perhaps with modifying phrases.</p> <p>D. The underlined words contain both independent clause(s) and dependent clause(s) perhaps with modifying phrases.</p>	17

relative pronouns (e.g., *who*, *whom*, *which*, etc.) or nominalizers (e.g., *that*, which may be understood).

Appendix B

Informed Consent Form

Dissertation Research: "Development of a Measure Assessing Knowledge and Use of Internal Punctuation to Signal Syntactic Relationships."

Thank you for accepting my invitation to participate in my study. What follows is very formal, but it is what I must disclose when conducting research.

You are invited to participate in a study that will assess your use of use of commas, semicolons, colons, phrases, and clauses. The study is conducted by Pamela Van Horn Howard to fulfill the dissertation requirements for a PhD in the Quantitative Research Methods Program at University of Denver. Results will be used to assess the reliability and validity of the measures created for the study to assess your performance on grammatical measures. Pamela Van Horn Howard can be reached at phoward2@du.edu. This project is supervised by Dissertation Advisor, Dr. Kathy Green, Department of Research Methods and Statistics, Morgridge College of Education, University of Denver, Denver, CO 80208 (303-871-2490/kgreen@du.edu).

Your consent to participate is highly valued. Participation in this study should take about 15 minutes of your time and will involve responding to 89 questions on a variety of grammar-related topics plus a very short writing sample. Participation in this project is strictly voluntary, and the risks associated with this project are minimal. If, however, you experience discomfort, you may discontinue participation at any time. Refusal to participate or withdrawal from participation will involve no penalty or loss of benefits to which you are otherwise entitled. However, to be eligible for the \$300 lottery, you must complete the entire questionnaire.

In order to protect the confidentiality anonymity of your input, your responses will be identified by code number only and will be kept separate from information that could identify you. Only the researcher will have access to your individual data.

If you have any concerns or complaints about matters related to your participation in this study, please contact Susan Sadler, Chair, Institutional Review Board for the Protection of Human Subjects, at 303-871-3454, or Sylk Sotto-Santiago, Office of Research and Sponsored Programs at 303-871-4052 or write to the University of Denver, Office of Research and Sponsored Programs, 2199 S. University Blvd., Denver, CO 80208-2121.

You may request an electronic copy of this informed consent form. If you have questions regarding any part of the above statement, please contact the researcher at phoward2@du.edu. Proceeding from this point, to complete the questionnaire will constitute your informed consent and testify to your understanding and agreement to the following:

I have read and understood the above descriptions of the study called "Development of a Measure Assessing Knowledge and Use of Internal Punctuation to Signal Syntactic Relationships." I have asked for and received a satisfactory explanation of any language that I did not fully understand. I agree to participate in this study, and I understand that I may withdraw my consent at any time. I have requested an electronic copy of this consent form if needed.

Appendix C

Syntactical Relationships as Signaled by Internal Punctuation: Multiple-Choice Grammar Test

I have always wondered just how well exercises and tests from books and worksheets actually reveal how much a student understands grammar. Because of that, I have developed a test below that is formatted in ways that I have not seen used before. I give you a chance later to tell me how good you think the test is.

First, please rate your confidence in using commas, semicolons and colons.

	Not at all confident	Somewhat confident	Confident	Very confident
Use of commas	-----	-----	-----	-----
Use of semicolons	-----	-----	-----	-----
Use of colons	-----	-----	-----	-----

Section 1: In this first section, you are asked to decide whether a comma, semicolon, or colon is *missing* from the sentence. For each of the following sentences, choose one answer that best clarifies the sentence's meaning from the choices given.

Example: This test may seem difficult but you will do well.

- A. Missing a comma
- B. Missing a semicolon
- C. Missing a colon
- D. No additional punctuation is needed.

1-1. We had been planning our ski trip for three months events didn't turn out as expected, however.

- A. Missing a comma
- B. Missing a semicolon
- C. Missing a colon
- D. No additional punctuation is needed.

1-2. Before we even made it onto the slopes we experienced our share of mini-disasters.

- A. Missing a comma
- B. Missing a semicolon
- C. Missing a colon
- D. No additional punctuation is needed.

1-3. First, the car wouldn't start and then the "check engine" light came on.

- A. Missing a comma
- B. Missing a semicolon
- C. Missing a colon
- D. No additional punctuation is needed.

1-4. To our amazement, however we found a small-town garage open for business at 6:00 a.m.

- A. Missing a comma
- B. Missing a semicolon
- C. Missing a colon
- D. No additional punctuation is needed.

1-5. Charging us \$400 the mechanic fixed our car.

- A. Missing a comma
- B. Missing a semicolon
- C. Missing a colon
- D. No additional punctuation is needed.

1-6. We arrived at the resort three hours behind schedule and found the parking lot completely full.

- A. Missing a comma
- B. Missing a semicolon
- C. Missing a colon
- D. No additional punctuation is needed.

1-7. We were redirected to an overflow lot where we found a spot in the very last row.

- A. Missing a comma
- B. Missing a semicolon
- C. Missing a colon
- D. No additional punctuation is needed.

1-8. We maneuvered into the spot just as the shuttle was pulling away.

- A. Missing a comma
- B. Missing a semicolon
- C. Missing a colon
- D. No additional punctuation is needed.

1-9. We had to wait in three long lines one for the shuttle, one for the tickets, and one for the ski lift.

- A. Missing a comma
- B. Missing a semicolon
- C. Missing a colon
- D. No additional punctuation is needed.

1-10. We managed two runs down the slopes which were very icy, and then I injured my ankle.

- A. Missing a comma
- B. Missing a semicolon
- C. Missing a colon
- D. No additional punctuation is needed.

1-11. We gave up and left for home.

- A. Missing a comma
- B. Missing a semicolon
- C. Missing a colon
- D. No additional punctuation is needed.

1-12. We learned this lesson from our experience leave for the slopes the day before you want to actually ski.

- A. Missing a comma
- B. Missing a semicolon
- C. Missing a colon
- D. No additional punctuation is needed.

Consider the questions in the section you have just completed. Please indicate by circling the vertical marker how well these types of questions assess your understanding of syntactical relationships with regard to punctuation.

Very poor Not very good Satisfactory Very good Excellent

|-----|-----|-----|-----|

Section 2: In this section, you are asked to decide whether you should *add* some punctuation.

Locate the caret (^) in the following sentences and choose the answer that best clarifies the sentence's meaning from the choices given.

Example: This test may seem difficult but you will do well.

^

- A. Insert a comma.
- B. Insert a semicolon.
- C. Insert a colon.
- D. No additional punctuation is needed

2-1. Because we had been warned of the impending snowstorm, we bought extra food and set the snow shovel by the front door.

^

- A. Insert a comma.
- B. Insert a semicolon.
- C. Insert a colon.
- D. No additional punctuation is needed.

2-2. The windows were drafty so we applied some weather-stripping.

^

- A. Insert a comma.
- B. Insert a semicolon.
- C. Insert a colon.
- D. No additional punctuation is needed.

2-3. Before the temperature dropped we cut some firewood for the wood-burning stove.

^

- A. Insert a comma.
- B. Insert a semicolon.
- C. Insert a colon.
- D. No additional punctuation is needed.

2-4. We lit a fire as the snow began to fall.

^

- A. Insert a comma.
- B. Insert a semicolon.
- C. Insert a colon.
- D. No additional punctuation is needed.

2-5. We drank hot cocoa by the fire which was casting an amber glow throughout the room.

^

- A. Insert a comma.
- B. Insert a semicolon.
- C. Insert a colon.
- D. No additional punctuation is needed.

2-6. We were delighted to see the snow piling up maybe school would be cancelled.

^

- A. Insert a comma.
- B. Insert a semicolon.
- C. Insert a colon.
- D. No additional punctuation is needed.

2-7. We burned all of the wood that we had chopped earlier.

^

- A. Insert a comma.
- B. Insert a semicolon.
- C. Insert a colon.
- D. No additional punctuation is needed.

2-8. We awoke to a foot of snow and news that the city was closed for business.

^

- A. Insert a comma.
- B. Insert a semicolon.
- C. Insert a colon.
- D. No additional punctuation is needed.

2-9. My favorite activities on snow days are sleeping in, reading a good book, and taking a walk outside.

^

- A. Insert a comma.
- B. Insert a semicolon.
- C. Insert a colon.
- D. No additional punctuation is needed.

2-10. Instead of those activities however, I usually have to shovel the walk, unbury the car, and make sure the pipes don't freeze.

^

- A. Insert a comma.
- B. Insert a semicolon.
- C. Insert a colon.
- D. No additional punctuation is needed.

2-11. In all honesty I hope the snow never melts.

^

- A. Insert a comma.
- B. Insert a semicolon.
- C. Insert a colon.
- D. No additional punctuation is needed.

Consider the questions in the section you have just completed. Please indicate by circling the vertical marker how well these types of questions assess your understanding of syntactical relationships with regard to punctuation.

Very poor Not very good Satisfactory Very good Excellent

|-----|-----|-----|-----|

Section 3: In this section, you are asked to decide whether you should *retain*, *remove*, or *replace* commas. For each of the following sentences, choose one answer that best clarifies the sentence’s meaning from the choices given.

Example: This test may seem difficult, but you will do well.

- A. Retain all commas.
- B. Remove all commas.
- C. Replace comma(s) with semicolon(s).

3-1. Anyone, who has an interest in photography, is encouraged to sign up for Journalism 101 this semester.

- A. Retain all commas.
- B. Remove all commas.
- C. Replace comma(s) with semicolon(s).

3-2. A journalistic photographer will work with the student news writers, and will be required to photograph events on campus as needed.

- A. Retain all commas.
- B. Remove all commas.
- C. Replace comma(s) with semicolon(s).

3-3. The duties of a journalistic photographer are challenging, the hours are flexible though.

- A. Retain all commas.
- B. Remove all commas.
- C. Replace comma(s) with semicolon(s).

3-4. The journalism department partners with the film department to teach photographers how to process film, which presents added benefits to staff photographers.

- A. Retain all commas.
- B. Remove all commas.
- C. Replace comma(s) with semicolon(s).

3-5. Each class admits only two student photographers, because space is limited.

- A. Retain all commas.
- B. Remove all commas.
- C. Replace comma(s) with semicolon(s).

3-6. The journalism class photographers work under the guidance of paid staff photographers of the student newspaper, paid photographers have already completed the journalism class.

- A. Retain all commas.
- B. Remove all commas.
- C. Replace comma(s) with semicolon(s).

3-7. In general, the journalism course offers young journalists both an education in journalistic techniques and practical writing experience.

- A. Retain all commas.
- B. Remove all commas.
- C. Replace comma(s) with semicolon(s).

3-8. Because the class is so popular, applicants for the course must submit a writing sample that conforms to AP Publication Style.

- A. Retain all commas.
- B. Remove all commas.
- C. Replace comma(s) with semicolon(s).

3-9. Style manuals for AP, the preferred publication style for most newspapers, may be purchased at the bookstore.

- A. Retain all commas.
- B. Remove all commas.
- C. Replace comma(s) with semicolon(s).

3-10. The AP manual is reasonably priced, and it is one book that journalists are sure to use often.

- A. Retain all commas.
- B. Remove all commas.
- C. Replace comma(s) with semicolon(s).

3-11. The class is demanding, it takes hard work and long hours to produce a quality student newspaper.

A. Retain all commas.

B. Remove all commas.

C. Replace comma(s) with semicolon(s).

Consider the questions in the section you have just completed. Please indicate how well these types of questions test your understanding of commas, semicolons, and colons.

Very poor Not very good Satisfactory Very good Excellent

|-----|-----|-----|-----|

Regarding your knowledge of commas, semicolons, and colons, please indicate how much of your understanding is governed by either intuition or direct instruction.

My understanding

is purely
intuitive

is more intuitive
than from direct
instruction

is an even mixture
of these extremes

is derived more
from direct
instruction than
from intuition

is derived solely from
direct instruction

|-----|-----|-----|-----|

Section 4: In this section, the focus shifts from internal punctuation to phrases and clauses. You will be asked to identify whether underlined passages are phrases or clauses.

First, please rate your confidence in your knowledge of phrases and clauses.

Not at all confident Somewhat confident Confident Very confident

Your knowledge of phrases |-----|-----|-----|

Your knowledge of clauses |-----|-----|-----|

In each sentence below, some of the words are underlined. Circle the answer that best reflects your understanding of phrases and clauses. Here are some definitions to jog your memory:

Phrase is defined as a unit of words that does not contain a subject-verb relationship.

Independent Clause is defined as a unit of words that does contain a subject-verb relationship and can stand alone. Clauses remain independent if they are joined by coordinating conjunctions (e.g., *and, or, but*, etc.) or conjunctive adverbs (e.g., *however, furthermore, therefore*, etc.).

Dependent clause is defined as a unit of words that does contain a subject-verb relationship yet cannot stand alone. Clauses most often are made dependent when paired with subordinating conjunctions (e.g., *because, while*, etc.), relative pronouns (e.g., *who, whom, which*, etc.) or nominalizers (e.g., *that*, which may be understood without actually appearing in the sentence—I knew *(that)* the flashdrive was in my backpack.).

Example: This test may seem difficult, but you will do well.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases..

Here is the story first:

I knew that I could miss my flight if I did not finish packing quickly. Once I threw my luggage into the trunk, started the car, and backed down the driveway, I realized I had forgotten to pack the gift that I intended to give my host, an oversight setting me back even further. As I sped down the highway on my way to the airport, I passed a state patrolman parked just beyond the crest of a hill, out of sight until it was too late for me to slow down. Of course, she handed me a speeding ticket, remarking as she walked away that all the flights were delayed owing to a breach in airport security. Thankfully, I arrived at the airport in time to catch my plane. I gave my keys to the valet and watched him drive away. At that moment, I realized I had forgotten my cell phone in the car, but it was too late to retrieve it unless I wanted to miss the plane after all. Why do I seem to procrastinate when my time is most limited?

4-1. I knew that I could miss my flight if I did not finish packing quickly.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-2. Once I threw my luggage into the trunk, started the car, and backed down the driveway, I realized I had forgotten to pack the gift that I intended to give my host, an oversight setting me back even further.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-3. Once I threw my luggage into the trunk, started the car, and backed down the driveway, I realized I had forgotten to pack the gift that I intended to give my host, an oversight setting me back even further.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-4. Once I threw my luggage into the trunk, started the car, and backed down the driveway, I realized I had forgotten to pack the gift that I intended to give my host, an oversight setting me back even further.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-5. Once I threw my luggage into the trunk, started the car, and backed down the driveway, I realized I had forgotten to pack the gift that I intended to give my host, an oversight setting me back even further.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-6. As I sped down the highway on my way to the airport, I passed a state patrolman parked just beyond the crest of a hill, out of sight until it was too late for me to slow down.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-7. As I sped down the highway on my way to the airport, I passed a state patrolman parked just beyond the crest of a hill, out of sight until it was too late for me to slow down.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-8. As I sped down the highway on my way to the airport, I passed a state patrolman parked just beyond the crest of a hill, out of sight until it was too late for me to slow down.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-9. As I sped down the highway on my way to the airport, I passed a state patrolman parked just beyond the crest of a hill, out of sight until it was too late for me to slow down.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-10. Of course, she handed me a speeding ticket, remarking as she walked away that all the flights were delayed owing to a breach in airport security.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-11. Of course, she handed me a speeding ticket, remarking as she walked away that all the flights were delayed owing to a breach in airport security.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-12. Of course, she handed me a speeding ticket, remarking as she walked away that all the flights were delayed owing to a breach in airport security.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-13. Thankfully, I arrived at the airport in time to catch my plane.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-14. I gave my keys to the valet and watched him drive away.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-15. At that moment, I realized I had forgotten my cell phone in the car, but it was too late to retrieve it unless I wanted to miss the plane after all.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-16. At that moment, I realized I had forgotten my cell phone in the car, but it was too late to retrieve it unless I wanted to miss the plane after all.

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

4-17. Why do I seem to procrastinate when my time is most limited?

The underlined words constitute

- A. **phrase(s)** only.
- B. **an independent clause** perhaps with modifying phrases.
- C. **a dependent clause** perhaps with modifying phrases.
- D. **both independent clause(s) and dependent clause(s)** perhaps with modifying phrases.

Consider the questions in the section you have just completed. How well do these types of questions test your knowledge of phrases and clauses?

Very poor Not very good Satisfactory Very good Excellent

|-----|-----|-----|-----|

Regarding your knowledge of phrases and clauses, please indicate how much of your understanding is governed by either intuition or direct instruction.

My understanding

is purely intuitive

is more intuitive than from direct instruction

is an even mixture of these extremes

is derived more from direct instruction than from intuition

is derived solely from direct instruction



Self Assessment: Understanding of Sentence Structures

Please indicate your level of agreement with the following statements regarding sentence structure and internal punctuation.

Understanding of Sentence Structures

	Rarely	Sometimes	About half the time	More often than not	Almost always
I can recognize an independent, or main, clause .	-----	-----	-----	-----	-----
a dependent, or subordinate, clause .	-----	-----	-----	-----	-----
a phrase .	-----	-----	-----	-----	-----
a fragmented sentence .	-----	-----	-----	-----	-----
a run on sentence .	-----	-----	-----	-----	-----
a comma splice .	-----	-----	-----	-----	-----

When teachers point out internal punctuation use in my own writing instead of in someone else's writing, I understand how to punctuate more effectively.

|-----|-----|-----|-----|

It is normal for my teachers to ask me to talk about why I punctuate as I do.

|-----|-----|-----|-----|

State of Student's Own Learning: Sentence Structure and Internal Punctuation

Strongly Disagree Disagree Agree Strongly Agree

I am confused about how to use internal punctuation. *

|-----|-----|-----|

I would use internal punctuation more consistently if I understood more clearly how to use it. *

|-----|-----|-----|

I need more help in learning how to use internal punctuation. *

|-----|-----|-----|

I am criticized for my use of internal punctuation. *

|-----|-----|-----|

I can tell you why I use internal punctuation as I do.

|-----|-----|-----|

I like learning how to use internal punctuation.

|-----|-----|-----|

I use internal punctuation correctly.

|-----|-----|-----|

I have learned how to use internal punctuation so that my readers understand the relationships among my ideas in the sentence.

|-----|-----|-----|

(*For data entry, reverse score)

Writing Sample

Please write 5-7 sentences regarding the grammar instruction you have received. Maybe you really enjoyed learning grammar; maybe you did not enjoy it very much but knew it was important to learn; maybe you did not feel that instruction was useful or consistent enough for you to learn it to your satisfaction. In addition, you may remember certain strategies or materials that your instructors used to teach you grammar. Talk about their effectiveness. How does your confidence in your own grammar use impact your writing now, especially in your profession? If you have a story that is pivotal in your experience with grammar use or instruction, please tell it.

PLEASE WRITE 5-7 SENTENCES

Please provide the following demographic information:

What type of high school did you attend? (Check all that apply.)

Public ___ Private (nonreligious) ___ Private (religious) ___ Home School ___ Alternative or Charter ___ GED ___

What is your level of education?

Some high school ___ High School graduate ___ Some college ___ Two-year or trade school degree ___ Bachelor's degree ___
Master's degree ___ Beyond Master's ___ Other ___ (please specify)

Are you a student in a degree-seeking program?

If you are a student

What is the name of your institution?

What is the name of your program?

What is your profession?

How important is writing in your daily life?

In your home life	Not important	Somewhat important	Important	Essential	n/a
-------------------	---------------	--------------------	-----------	-----------	-----

Your professional life	Not important	Somewhat important	Important	Essential	n/a
------------------------	---------------	--------------------	-----------	-----------	-----

Your academic life	Not important	Somewhat important	Important	Essential	n/a
--------------------	---------------	--------------------	-----------	-----------	-----

When you do write, do you feel that others will scrutinize your grammar use?

Rarely sometimes about half the time more often than not almost always

Do you scrutinize your grammar use?

Rarely sometimes about half the time more often than not almost always

Gender: _____

Race/Ethnicity: _____

Year of high school graduation or GED _____

Age: _____

Average High School Language Arts Grade: A B C D F Other

Please follow the link to enter the \$300 lottery.

Appendix D

All SRSIP Items in Misfit Order: Winsteps Output

INPUT: 328 PERSONS 51 items MEASURED: 320 PERSONS 51 items 2 CATS 1.0.0

PERSON: REAL SEP.: 1.80 REL.: .76 ... ITEM: REAL SEP.: 8.17 REL.: .99

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ ZSTD	OUTFIT MNSQ ZSTD	PTMEA CORR.	EXACT OBS%	MATCH EXP%	ITEM
37	35	273	3.40	.19	1.09 .8	2.32 4.8	A .05	87.9	87.5	4_3RO
8	292	320	-1.48	.21	1.04 .3	1.79 2.6	B .21	90.9	91.2	1_8RF
23	37	313	3.53	.18	1.13 1.0	1.46 2.0	C .11	87.2	88.5	2_2RD
2	221	320	.26	.13	1.17 2.8	1.26 2.7	D .19	67.8	72.9	1_2REG
3	158	320	1.24	.12	1.12 2.7	1.22 3.5	E .26	62.8	66.1	1_3RD
49	71	273	2.42	.15	1.11 1.5	1.21 1.8	F .21	73.6	75.6	4_15RO
7	222	320	.25	.13	1.12 2.1	1.20 2.1	G .24	70.0	73.1	1_7RIJ
51	93	273	1.98	.14	1.08 1.5	1.18 2.0	H .26	69.2	70.0	4_17RO
22	278	313	-1.16	.19	.99 -.1	1.14 .7	I .26	89.1	88.9	2_11REG
20	105	313	2.02	.13	1.01 .3	1.13 1.6	J .33	70.9	70.7	2_9RK
50	95	273	1.95	.14	1.05 1.0	1.12 1.4	K .29	68.1	69.6	4_16RM
16	232	313	-.01	.14	1.04 .6	1.10 .9	L .30	74.4	76.1	2_5RIJ
48	144	273	1.09	.13	1.08 1.9	1.08 1.3	M .29	64.1	66.1	4_14RM
39	119	273	1.52	.13	1.08 1.8	1.08 1.2	N .29	60.8	66.3	4_5RL
14	246	313	-.29	.15	1.08 1.0	1.08 .6	O .26	78.0	79.7	2_3REG
27	250	304	-.54	.16	.97 -.3	1.07 .5	P .33	83.6	82.8	3_4RIJ
6	283	320	-1.14	.18	1.00 .0	1.06 .4	Q .27	88.1	88.5	1_6RC
28	224	304	.03	.14	1.03 .4	1.06 .6	R .32	76.0	75.7	3_5RF
21	230	313	.03	.14	1.02 .3	1.06 .6	S .33	75.1	75.6	2_10RH
33	165	304	1.03	.12	1.05 1.1	1.05 .8	T .33	62.8	66.4	3_10RD
47	165	273	.72	.13	1.03 .6	1.04 .5	U .34	67.8	68.3	4_13RM
25	174	304	.89	.13	1.02 .4	1.00 .0	V .36	65.8	67.1	3_2RC
11	299	320	-1.81	.23	1.02 .2	.94 -.1	W .20	93.4	93.4	1_11RC
43	154	273	.91	.13	1.01 .3	1.01 .1	X .36	65.9	67.0	4_9RN
38	170	273	.63	.13	1.01 .2	1.00 .1	Y .36	69.2	69.0	4_4RN
45	157	273	.86	.13	1.00 .1	.97 -.4	Y .38	66.3	67.3	4_11RN
36	162	273	.77	.13	1.00 -.1	.98 -.2	x .38	65.9	67.9	4_2RN
10	223	320	.23	.13	.99 -.2	.99 -.1	w .38	73.8	73.3	1_10IJ
42	146	273	1.05	.13	.98 -.3	.95 -.8	v .40	63.7	66.3	4_8RL
17	270	313	-.90	.17	.98 -.1	.80 -1.1	u .34	85.6	86.5	2_6RAB
15	299	313	-2.22	.28	.98 .0	.67 -.9	t .25	95.5	95.5	2_4RF
5	286	320	-1.24	.19	.97 -.1	.72 -1.3	s .34	89.7	89.4	1_5REG
46	169	273	.65	.13	.96 -.8	.92 -1.0	r .42	71.8	68.9	4_12RL
35	33	273	3.47	.19	.96 -.3	.92 -.3	q .31	88.6	88.2	4_1RO
24	265	304	-.97	.18	.95 -.4	.78 -1.1	p .36	87.2	87.4	3_1RIJ
9	244	320	-.16	.14	.95 -.7	.85 -1.3	o .42	77.2	77.9	1_9RK
12	235	320	.02	.14	.93 -1.0	.92 -.7	n .43	77.2	75.8	1_12RK
30	299	304	-3.27	.46	.93 .0	.33 -1.4	m .24	98.4	98.4	3_7REG
26	264	304	-.94	.18	.92 -.6	.92 -.4	l .36	87.5	87.0	3_3RAB
19	298	313	-2.14	.27	.92 -.3	.62 -1.1	k .30	95.2	95.2	2_8RC
4	276	320	-.92	.17	.92 -.7	.90 -.5	j .36	87.5	86.4	1_4RH
31	292	304	-2.34	.30	.92 -.2	.49 -1.5	i .30	96.1	96.0	3_8REG
18	308	313	-3.31	.45	.90 -.1	.26 -1.6	h .27	98.4	98.4	2_7RIJ
32	292	304	-2.34	.30	.89 -.4	.44 -1.7	g .33	96.1	96.0	3_9RH
41	215	273	-.30	.16	.86 -1.6	.68 -2.6	f .50	79.1	79.7	4_7RL
29	287	304	-1.95	.26	.85 -.7	.51 -1.7	e .38	94.4	94.4	3_6RAB
1	282	320	-1.10	.18	.85 -1.2	.71 -1.4	d .43	88.4	88.2	1_1RAB
40	164	273	.74	.13	.85 -3.2	.79 -3.1	c .54	75.5	68.1	4_6RN
44	221	273	-.46	.16	.84 -1.7	.67 -2.4	b .50	81.7	81.6	4_10RM
34	256	304	-.70	.17	.83 -1.7	.59 -2.7	a .50	85.5	84.6	3_11RAB
MEAN	204.0	299.4	.17	.21	.99 .1	.96 .0		79.4	79.9	
S.D.	81.7	19.4	1.98	.24	.08 1.1	.34 1.6		11.2	10.5	

Appendix E

SRSIP Punctuation Items Items in Misfit Order: Winsteps Output

INPUT: 328 PERSONS 51 items MEASURED: 320 PERSONS 34 items 2 CATS 1.0.0

PERSON: REAL SEP.: 1.80 REL.: .76 ... ITEM: REAL SEP.: 8.17 REL.: .99

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEA CORR.	EXACT OBS%	MATCH EXP%	ITEM
					MNSQ	ZSTD	MNSQ	ZSTD				
8	292	320	-1.48	.21	1.04	.3	1.79	2.6	A .21	90.9	91.2	1_8RF
23	37	313	3.53	.18	1.13	1.0	1.46	2.0	B .11	87.2	88.5	2_2RD
2	221	320	.26	.13	1.17	2.8	1.26	2.7	C .19	67.8	72.9	1_2REG
3	158	320	1.24	.12	1.12	2.7	1.22	3.5	D .26	62.8	66.1	1_3RD
7	222	320	.25	.13	1.12	2.1	1.20	2.1	E .24	70.0	73.1	1_7RIJ
22	278	313	-1.16	.19	.99	-.1	1.14	.7	F .26	89.1	88.9	2_11REG
20	105	313	2.02	.13	1.01	.3	1.13	1.6	G .33	70.9	70.7	2_9RK
16	232	313	-.01	.14	1.04	.6	1.10	.9	H .30	74.4	76.1	2_5RIJ
14	246	313	-.29	.15	1.08	1.0	1.08	.6	I .26	78.0	79.7	2_3REG
27	250	304	-.54	.16	.97	-.3	1.07	.5	J .33	83.6	82.8	3_4RIJ
6	283	320	-1.14	.18	1.00	.0	1.06	.4	K .27	88.1	88.5	1_6RC
28	224	304	.03	.14	1.03	.4	1.06	.6	L .32	76.0	75.7	3_5RF
21	230	313	.03	.14	1.02	.3	1.06	.6	M .33	75.1	75.6	2_10RH
33	165	304	1.03	.12	1.05	1.1	1.05	.8	N .33	62.8	66.4	3_10RD
25	174	304	.89	.13	1.02	.4	1.00	.0	O .36	65.8	67.1	3_2RC
11	299	320	-1.81	.23	1.02	.2	.94	-.1	P .20	93.4	93.4	1_11RC
10	223	320	.23	.13	.99	-.2	.99	-.1	Q .38	73.8	73.3	1_10IJ
17	270	313	-.90	.17	.98	-.1	.80	-1.1	p .34	85.6	86.5	2_6RAB
15	299	313	-2.22	.28	.98	.0	.67	-.9	o .25	95.5	95.5	2_4RF
5	286	320	-1.24	.19	.97	-.1	.72	-1.3	n .34	89.7	89.4	1_5REG
24	265	304	-.97	.18	.95	-.4	.78	-1.1	m .36	87.2	87.4	3_1RIJ
9	244	320	-.16	.14	.95	-.7	.85	-1.3	l .42	77.2	77.9	1_9RK
12	235	320	.02	.14	.93	-1.0	.92	-.7	k .43	77.2	75.8	1_12RK
30	299	304	-3.27	.46	.93	.0	.33	-1.4	j .24	98.4	98.4	3_7REG
26	264	304	-.94	.18	.92	-.6	.92	-.4	i .36	87.5	87.0	3_3RAB
19	298	313	-2.14	.27	.92	-.3	.62	-1.1	h .30	95.2	95.2	2_8RC
4	276	320	-.92	.17	.92	-.7	.90	-.5	g .36	87.5	86.4	1_4RH
31	292	304	-2.34	.30	.92	-.2	.49	-1.5	f .30	96.1	96.0	3_8REG
18	308	313	-3.31	.45	.90	-.1	.26	-1.6	e .27	98.4	98.4	2_7RIJ
32	292	304	-2.34	.30	.89	-.4	.44	-1.7	d .33	96.1	96.0	3_9RH
29	287	304	-1.95	.26	.85	-.7	.51	-1.7	c .38	94.4	94.4	3_6RAB
1	282	320	-1.10	.18	.85	-1.2	.71	-1.4	b .43	88.4	88.2	1_1RAB
34	256	304	-.70	.17	.83	-1.7	.59	-2.7	a .50	85.5	84.6	3_11RAB
MEAN	238.0	312.6	-.38	.24	.98	.1	.91	.0		83.3	83.9	
S.D.	71.7	6.6	2.11	.29	.08	1.0	.32	1.5		10.5	9.7	

Appendix F

Statistically Significant Results of Punctuation Items Differential Item Functioning Based upon Education Level: Winsteps Output

INPUT: 328 PERSONS 51 fullsets MEASURED: 320 PERSONS 34 fullsets 2 CATS 1.0.0

Statistical significance = $p < .01$

PERSON	DIF	DIF	PERSON	DIF	DIF	DIF	JOINT	MantelHanzl fullset						
CLASS	MEASURE	S.E.	CLASS	MEASURE	S.E.	CONTRAST	S.E.	t	d.f.	Prob.	Prob.	Size	Number	Name
3	-.25	.27	6	.79	.25	-1.04	.37	-2.79	153	.0060	.0597	-.66	2	1_2REG
3	.91	.25	5	1.92	.27	-1.01	.37	-2.77	142	.0064	.0123	.62	3	1_3RD
2	.95	1.18	3	4.81	.72	-3.86	1.39	-2.79	80	.0067	.3173	-	23	2_2RD

Size of Mantel-Haenszel slice = .100 logits

Appendix G

SRSIP Phrase/Clause Identification Items in Misfit Order: Winsteps Output

INPUT: 328 PERSONS 51 items MEASURED: 320 PERSONS 17 items 2 CATS 1.0.0

PERSON: REAL SEP.: 1.80 REL.: .76 ... ITEM: REAL SEP.: 8.17 REL.: .99

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEA CORR.	EXACT OBS%	MATCH EXP%	ITEM
37	35	273	3.40	.19	1.09	.8	2.32	4.8	A .05	87.9	87.5	4_3RO
49	71	273	2.42	.15	1.11	1.5	1.21	1.8	B .21	73.6	75.6	4_15RO
51	93	273	1.98	.14	1.08	1.5	1.18	2.0	C .26	69.2	70.0	4_17RO
50	95	273	1.95	.14	1.05	1.0	1.12	1.4	D .29	68.1	69.6	4_16RM
48	144	273	1.09	.13	1.08	1.9	1.08	1.3	E .29	64.1	66.1	4_14RM
39	119	273	1.52	.13	1.08	1.8	1.08	1.2	F .29	60.8	66.3	4_5RL
47	165	273	.72	.13	1.03	.6	1.04	.5	G .34	67.8	68.3	4_13RM
43	154	273	.91	.13	1.01	.3	1.01	.1	H .36	65.9	67.0	4_9RN
38	170	273	.63	.13	1.01	.2	1.00	.1	I .36	69.2	69.0	4_4RN
45	157	273	.86	.13	1.00	.1	.97	-.4	h .38	66.3	67.3	4_11RN
36	162	273	.77	.13	1.00	-.1	.98	-.2	g .38	65.9	67.9	4_2RN
42	146	273	1.05	.13	.98	-.3	.95	-.8	f .40	63.7	66.3	4_8RL
46	169	273	.65	.13	.96	-.8	.92	-1.0	e .42	71.8	68.9	4_12RL
35	33	273	3.47	.19	.96	-.3	.92	-.3	d .31	88.6	88.2	4_1RO
41	215	273	-.30	.16	.86	-1.6	.68	-2.6	c .50	79.1	79.7	4_7RL
40	164	273	.74	.13	.85	-3.2	.79	-3.1	b .54	75.5	68.1	4_6RN
44	221	273	-.46	.16	.84	-1.7	.67	-2.4	a .50	81.7	81.6	4_10RM

MEAN	136.1	273.0	1.26	.14	1.00	.1	1.05	.1		71.7	72.2	
S.D.	53.0	.0	1.06	.02	.08	1.3	.35	1.9		8.0	7.2	

SRSIP Phrase/Clause Identification Items in Misfit Order with Item 4-3 Omitted: Winsteps Output

INPUT: 328 PERSONS 51 items MEASURED: 320 PERSONS 16 items 2 CATS 1.0.0

PERSON: REAL SEP.: 1.80 REL.: .76 ... ITEM: REAL SEP.: 8.17 REL.: .99

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEA CORR.	EXACT OBS%	MATCH EXP%	ITEM
49	71	273	2.42	.15	1.11	1.5	1.21	1.8	A .21	73.6	75.6	4_15RO
51	93	273	1.98	.14	1.08	1.5	1.18	2.0	B .26	69.2	70.0	4_17RO
50	95	273	1.95	.14	1.05	1.0	1.12	1.4	C .29	68.1	69.6	4_16RM
48	144	273	1.09	.13	1.08	1.9	1.08	1.3	D .29	64.1	66.1	4_14RM
39	119	273	1.52	.13	1.08	1.8	1.08	1.2	E .29	60.8	66.3	4_5RL
47	165	273	.72	.13	1.03	.6	1.04	.5	F .34	67.8	68.3	4_13RM
43	154	273	.91	.13	1.01	.3	1.01	.1	G .36	65.9	67.0	4_9RN
38	170	273	.63	.13	1.01	.2	1.00	.1	H .36	69.2	69.0	4_4RN
45	157	273	.86	.13	1.00	.1	.97	-.4	h .38	66.3	67.3	4_11RN
36	162	273	.77	.13	1.00	-.1	.98	-.2	g .38	65.9	67.9	4_2RN
42	146	273	1.05	.13	.98	-.3	.95	-.8	f .40	63.7	66.3	4_8RL
46	169	273	.65	.13	.96	-.8	.92	-1.0	e .42	71.8	68.9	4_12RL
35	33	273	3.47	.19	.96	-.3	.92	-.3	d .31	88.6	88.2	4_1RO
41	215	273	-.30	.16	.86	-1.6	.68	-2.6	c .50	79.1	79.7	4_7RL
40	164	273	.74	.13	.85	-3.2	.79	-3.1	b .54	75.5	68.1	4_6RN
44	221	273	-.46	.16	.84	-1.7	.67	-2.4	a .50	81.7	81.6	4_10RM

MEAN	142.4	273.0	1.12	.14	1.00	.1	.98	-.1		70.7	71.2	
S.D.	48.1	.0	.95	.02	.08	1.3	.15	1.5		7.2	6.3	

Appendix H

Statistically Significant Results of Phrase/Clause Identification Items (Minus Item 4-3) Differential Item Analysis Based upon Education Level: Winsteps Output

INPUT: 328 PERSONS 51 fullsets MEASURED: 320 PERSONS 16 fullsets 2 CATS 1.0.0

 Statistical significance = p < .01

PERSON CLASS	DIF MEASURE	DIF S.E.	PERSON CLASS	DIF MEASURE	DIF S.E.	DIF CONTRAST	JOINT S.E.	t	d.f.	Prob.	MantelHanzl Prob.	fullset Size	Number	Name
3	3.01	.35	6	4.77	.53	-1.76	.64	-2.75	153	.0067	.0738	-.63	35	4_1RO
5	3.05	.33	6	4.77	.53	-1.73	.63	-2.74	141	.0070	.0129	-	35	4_1RO
3	2.68	.32	5	1.49	.26	1.19	.41	2.87	142	.0047	.0005	1.47	50	4_16RM
4	-.53	.84	3	2.15	.28	-2.68	.89	-3.03	84	.0033	.0011	-	51	4_17RO
4	-.53	.84	5	2.22	.28	-2.75	.89	-3.10	72	.0027	.0630	-2.30	51	4_17RO
4	-.53	.84	6	1.83	.25	-2.36	.88	-2.69	83	.0087	.8084	+	51	4_17RO
4	-.53	.84	7	2.11	.35	-2.64	.91	-2.89	46	.0058	.9160	.69	51	4_17RO

 Size of Mantel-Haenszel slice = .100 logits

Appendix I

Means and Standard Deviations for SRSIP Test Section Person Logit Scores, Survey Items, and Writing Sample Error Rate

Scale or Item	Possible Score	Mean	SD	Skewness	Kurtosis	n
Punctuation Items Person Logit Scores	na	1.88	1.12	-.62	-.17	328
Phrase/Clause Identification Items (no item 4-3) Person Logit Scores	na	.09	1.24	-.01	-1.81	273
Confidence Item Total	20	13.27	3.43	-.01	-.64	304
Confidence Punctuation Item Total	12	8.88	2.34	-.24	-.83	327
Confidence Sentence Structure Total	8	4.36	1.6	.33	-.35	304
Confidence Comma Use	4	3.20	.77	-.57	-.42	327
Confidence Semicolon Use	4	2.83	.92	-.20	-.96	327
Confidence Colon Use	4	2.85	.96	-.33	-.92	327
Confidence Phrase Use	4	2.22	.81	.27	-.38	304
Confidence Clause Use	4	2.14	.83	.37	-.37	304
Sentence Structure Item Total	50	35.97	8.91	-.56	-.10	273
Sentence Structure Punctuation Total	35	26.16	6.17	-.76	.52	273
Sentence Structure Components Total	15	9.81	3.47	-.41	-.77	273
Recognize Independent Clause	5	3.28	1.31	-.40	-1.05	273
Recognize Dependent Clause	5	3.08	1.26	-.22	-1.05	273
Recognize Phrase	5	3.45	1.24	-.46	-1.56	273
Recognize Run on Sentence	5	4.25	1.01	-1.60	2.20	273
Recognize Comma Splice	5	3.29	1.31	-.36	-.95	273
Know Punctuation between Two Independent Clauses	5	3.43	1.27	-.50	-.84	273
Know Punctuation between Independent and Dependent Clauses	5	3.35	1.19	-.46	-.77	273
Know when to Use Commas	5	4.11	.85	-1.22	2.23	273
Know when to Use Semicolons	5	3.88	1.05	-.77	-.03	273
Know when to Use Colons	5	3.86	1.13	-.92	.20	273
Writing Sample Error Rate	1.0	.255	.241	1.20	1.35	328

Appendix J

Bivariate Correlations and Coefficients among Validating Survey Items and Error Rate

Scale or Item	Validating Correlations and Coefficients
Writing Sample Error Rate	Confidence Item Total: -.11 Sentence Structure Item Total: -.15*
Confidence Punctuation Item Total	Sentence Structure Punctuation Total: .72** Confidence Sentence Structure Total: .52**
Confidence Sentence Structure Total	Sentence Structure Components Total: .59**
Confidence Comma Use	Confidence Semicolon Use: .61** Confidence Colon Use: .62** Confidence Phrase Use: .40** Confidence Clause Use: .41**
Confidence Semicolon Use	Confidence Colon Use: .77** Confidence Phrase Use: .45** Confidence Clause Use: .49**
Confidence Colon Use	Confidence Phrase Use: .48** Confidence Clause Use: .50**
Confidence Phrase Use	Confidence Clause Use: .91**
Confidence Clause Use	All other Confidence Items (see above)
Sentence Structure Punctuation Total	Sentence Structure Components Total: .68**
Recognize Independent Clause	Recognize Dependent Clause: .90** Recognize Phrase: .65** Recognize Run on Sentence: .34** Recognize Comma Splice: .45** Know Punctuation between Two Independent Clauses: .65** Know Punctuation between Independent and Dependent Clauses: .70** Know when to Use Commas: .37** Know when to Use Semicolons: .42** Know when to Use Colons: .42** Confidence Clause Use: .54**
Recognize Dependent Clause	Recognize Phrase: .68** Recognize Run on Sentence: .37** Recognize Comma Splice: .49** Know Punctuation between Two Independent Clauses: .70** Know Punctuation between Independent and Dependent Clauses: .68** Know when to Use Commas: .39** Know when to Use Semicolons: .45** Know when to Use Colons: .45** Confidence Clause Use: .54**
	Confidence Clause Use: .54**

Recognize Phrase	Recognize Run on Sentence: .47** Recognize Comma Splice: .39** Know Punctuation between Two Independent Clauses: .52** Know Punctuation between Independent and Dependent Clauses: .54** Know when to Use Commas: .41** Know when to Use Semicolons: .46** Know when to Use Colons: .43** Confidence Phrase Use: .53**
Recognize Run on Sentence	Recognize Comma Splice: .44** Know Punctuation between Two Independent Clauses: .41** Know Punctuation between Independent and Dependent Clauses: .41** Know when to Use Commas: .57** Know when to Use Semicolons: .58** Know when to Use Colons: .58**
Recognize Comma Splice	Know Punctuation between Two Independent Clauses: .47** Know Punctuation between Independent and Dependent Clauses: .45** Know when to Use Commas: .40** Know when to Use Semicolons: .45** Know when to Use Colons: .45**
Know Punctuation between Two Independent Clauses	Know Punctuation between Independent and Dependent Clauses: .91** Know when to Use Commas: .57** Know when to Use Semicolons: .59** Know when to Use Colons: .60** Confidence Clause Use: .53**
Know Punctuation between Independent and Dependent Clauses	Know when to Use Commas: .56** Know when to Use Semicolons: .60** Know when to Use Colons: .62** Confidence Clause Use: .55**
Know when to Use Commas	Know when to Use Semicolons: .70** Know when to Use Colons: .67** Confidence Comma Use: .63**
Know when to Use Semicolons	Know when to Use Colons: .86** Confidence Semicolon Use: .68**
Know when to Use Colons	All other Sentence Structure Items (see above) Confidence Colon Use: .66**

Note: ** correlations statistically significant at $p < .001$; * correlations statistically significant at $p < .05$