ABSTRACT

Individual Differences and Cognitive Complexity Investigated in Community College Writing

David E. Thomson, Ph.D.

Chairperson: A. Alexander Beaujean, Ph.D.

Synthesizing empirical findings about lifelong writing development, I tested a measurement of cognitive complexity (CC) toward understanding how the language of affect may interact with that language associated with acts of thinking. A community college in central Texas was the site of the study where I analyzed the essays of basic (n= 134) and advanced (n =89) composition students. Since vocational-track students make up about half of the enrollment, I compared those students' (n= 27) performance with traditional associate-degree seeking students (n =134). Additionally, I collected personality profiles from many of those students (n= 145) to explore any possible interaction of Neuroticism (N) on the affect component of the measure under investigation. Results showed small relations between CC and sex (*Cohen's* d =.24), CC and course level (d =.18), and CC and N (r =.1). Just as women tended to outscore men on CC, so did basic composition students in comparison to their advanced peers. There was almost no difference between vocational and traditional college track students. Overall, this study may present evidence of a natural-word-usage ceiling evident in the automated textual analysis software used to measure CC. More clearly it showed that CC as measured in the present study negatively correlates with standardized reading (r =-0.14) and

writing (r =-0.28) scores. I conclude by discussing the need to gather a broader lifespan sampling of whatever ability and trait characteristics detail CC as that realized in an academic community valuing the free and tolerant exchange of ideas.

Individual Differences and Cognitive Complexity Investigated in Community College Writing

by

David E. Thomson, B.A., M.F.A.

A Dissertation

Approved by the Department of Educational Psychology

Marley W. Watkins, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

A. Alexander Beaujean, Ph.D., Chairperson

Dennis A. Johnston, Ph.D.

Coretta M. Pittman, Ph.D.

Terrill F. Saxon, Ph.D.

Julie K. Ivey-Hatz, Ph.D.

Accepted by the Graduate School
August 2014

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I wish to gratefully acknowledge the many who have inspired, challenged and vetted this study. Among those too numerous to be named are the poets, blues librettists and story tellers who first wove me into the web of language. More recently, I credit Dr. A. Alexander Beaujean for helping me unite quantitative rigor with the endless archives of qualitative reflection, Dr. Terrill Saxon for still believing the history of educational psychology should form its newest practitioners (as well for as introducing me to the genius of Lev Vygotsky), Dr. Paul Illich for sharing his study insights and Dr. James Pennebaker, who brought a clinician's heart to care for all the words we use to frame our human experiences. Among these saints known and unknown are certainly a few great friends who set aside time to cook, talk, music and walk out the best of what I could imagine. In no particular order, I gratefully remember Dr. Marc Guidry and Monique Guidry, Dr. Scott and Jennifer Rushing, Robby Thigpen, Mark Long, Court and Vivian Atchinson, Dr. Jerry Park and Dr. Christina Chan-Park, Dr. Burt Burleson and Eric Howell, as well as as well as family members Mark and Carron Blodgett, Sean Thomson, James David and Lura Thomson, and David Gavin Thomson. Collectively, I couldn't have possibly found the Jacobean insistence on wrestling for a blessing had it not been for the faithful, open and fostering community of Dayspring Baptist Church. The last word on this reflection on researching the architectonics of language must be offered up to my wife and colleague, Dr. Cindy Wu, who shared her unqualified enthusiasm for this research line from its inception.

DEDICATION

To the many joined by their fascination with learning, writing and employing their skills for the commonweal

CHAPTER ONE

Introduction

In this chapter I state key terms, provide a statement of the problem investigated, detail the purpose of the present study and conclude with guiding research questions.

1.1 Key Terms

1.1.1 Individual Differences

The study of individual differences locates psychological phenomena such as intelligence and personality at the person-level rather than in the collective. Research on individual differences/differential psychology often spans three areas: (a) the development and use and psychological constructs (i.e., psychometrics); (b) the structure of cognitive ability, personality and other areas in which there is variability among individuals; and (c) the relationship of these differences to positive life outcomes (Anastasi, 1981; Deary, 2001). Fundamental to the study of individual differences was Spearman's (1904) discovery of *g*, or the grouping factor that summed both the high and positive correlations so frequently seen in achievement tests. *g* is not an individual attribute, but an inference of general ability that seems to link domains of intelligence to one another (i.e., vocabulary and arithmetic). While the field has since refined Spearman's insights (Cattell, 1946; Horn & Cattell, 1966) his emphases on the hierarchical nature of intelligence and trait stability remain central to differential psychology. Concurrent with much of the research in intelligence was the attempt to understand the individual through trait psychology. Allport and Odbert (1936) summed 4,000 traits as *cardinal, central* and *secondary*. A cardinal trait sums a lifetime of experience (i.e., relentless), while a central trait might be identified as a personal description (i.e., faithful) and a secondary trait

would be situational (i.e., nervous while waiting for a test result). Cattell (1946) reduced Allport's descriptive list to 171 closely related terms and from there deduced 16 traits (among them self-reliance, vigilance and perfectionism). Currently, the Five Factor Model (FFM) (McCrae & John, 1992) of personality identifies Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness as summative of the span of human differences in personality. McCrae and Costa (1986), however, cautioned that personality analysis merely situates the individual at a global level, or in relation to many others.

1.1.2 Cognitive Complexity

In describing how people mold themselves to their environments, Kelly (1955) first used the term *cognitive complexity* (CC). In the present study, I employ the term to describe the complex interactions of individual writers addressing topics to particular audiences. Anytime a writer writes, I maintain, some glimpse of that individual's history with reading and writing is revealed. In the present study, I expected cognitive complexity can be developed with college training. Accordingly, those with greater college exposure I expected to demonstrate greater levels of CC. In general, psychologists agreed that CC pulls upon information retrieval, schematization and analogy-making. For writing in particular, researchers located CC in word usage, clauses and punctuation (Hunt, 1964; Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007b; Sanders & Schilperoord, 2006). Most noted a hierarchical awareness framing CC. That is to claim that cognitively complex writers are aware of how one statement amplifies or reverses another. This hierarchical awareness both Spearman (1927) and Vygotsky (1934/1986) believed marked intellectual development.

1.1.3 Academic Writing

By academic writing I mean that writing produced particularly for a discipline, professional or generalist scholarly community. In tenor, I take academic writing to privilege reason over emotionality. Its tone is deliberative, tolerant and yet assertive. The academic writer, in presenting various perspectives on contentious issues, is necessarily complex. Badley (2009) described that deliberative complexity as a recursive process of establishing and then deconstructing concepts. Synthesizing multiple texts is one activity of academic writers. Another is engaging the long process of committing ideas to paper and revising until those ideas find a necessary audience (Cameron, Nairn, & Higgins, 2009). Moreover, academic writing may not only be discipline dependent, but also tacit(Lewis, 2010). That is to suggest that academic writing is not formulaic. It may only be mastered in consideration of a particular audience.

1.2 Statement of Problem

Assessment of writing is both an inconsistent and time-consuming endeavor (Cherry & Meyer, 1993). To the extent that a college composition course functions as a gateway course to facilitate students' attention to reason over emotionality, both the inconsistency and time expenditure of human grading could be ameliorated through automated text analysis. Unlike human graders, Graesser and McNamara (2012) noted, computers "provide instantaneous feedback, do not get fatigued, are consistent, are unbiased in assigning scores to particular individuals, provide greater detail on many dimensions, and can apply sophisticated algorithms that humans could never understand and apply" (p. 308). Great strides have been recently made in automated scoring systems in equating the holistic scores of humans with that done by computers. For example, automated scoring systems at present agree with human raters more than 80 percent of the time and score essays within one point 90 percent of the time (Graesser & McNamara, 2012). However,

similarity in holistic scoring does not solve the problem of quantitatively measuring levels of cognition as revealed in natural language usage; neither does it provide a measure of emotional self-regulation that could be useful in gauging competency within an academic community.

1.3 Purpose of Present Study

I employed an automated text analysis program to quantitatively measure cognition and its interaction with emotion. I did so, as well, to see if a measurement of cognitive complexity could detect individual learner differences. Toward doing so, I asked the following questions:

1. Is cognitive complexity (CC) in an individual's writing related to performance on standardized written language placement skills tests among post-secondary students?

2. Is post-secondary matriculation, as measured by degree progress, related to CC?

3. Is there a difference in CC between students enrolled in vocational curricula versus those who take a traditional core of general education classes?

4. Do males and females differ in their average level of CC?

5. Is CC related to the personality dimension known as Neuroticism (N)?

6. What is the strength of the interaction between N and sex?

CHAPTER TWO

Literature Review

2.1 The Social Context of Writing

Words reveal much about their users' thinking, present emotional experience, and even underlying psychological traits, such as personality and cognitive functioning. While natural language use is not frequently a variable that differential psychology examines, its uniqueness makes it ripe to separate one speaker or writer from another (i.e., as a marker of individual differences). Moreover, when we use language, we always do so in social context. Even if the language user is only thinking out loud when doing some new procedural task or making a to-do list, he or she still employs a common resource learned from others. Language use, then, is an event formed by the language user's (a) previous learning, (b) appeal to a particular audience, and (c) purpose for the text or utterance. This model of a dynamic, three-way process of language use is foundational to the composition-rhetoric field (Pfister & Petrik, 1980).

2.1.1 Writerly Complexity

The present study employed *writerly* as a necessary complexity qualifier and does so in acknowledgement of literary criticism detailing differences between written and spoken speech (Barthes, 1959/1972; Derrida, 1967/1976). *Writerly* implied those qualities unique to written speech. Among those features are lengthy non-essential clauses embedded within sentences. Writing, in particular, must be more complex than speech, for words seen on some surface may be more readily identifiable as objects in their own rite, whereas speech seems so ephemeral. In positing such, the present study acknowledged the late-career Heidegger, whose pri-

mary authorial output was literary criticism (Hofstadter, 1971). In his reflections on German-language poet Georg Trakl, Heidegger argued that poetic language separates itself from ornamental language by reconceptualizing the world. Words, in short, become objects in their own rite for the things they sign. In this gesture, Heidegger located his *Dasein*—or being present to one's particular moment—in language itself. Heidegger himself held "language belongs to the closest neighborhood of man's being"(Heidegger, 1950, p. 1121). The complexity Heidegger addressed, then, is both that of the individual fingerprinting sensibility in language use and that common resource legacied from one generation to the next. The present study was motivated by the attempt to capture both aspects of language use. I held that *Writerly* joined to *g*, or the general ability Spearman described (1904), would suggest that whatever *g* an individual possesses will be proportionately present as evidenced by a measure of cognitive complexity.

The scaling of one currently available text analysis software (Pennebaker, Chung, Ireland, Gonzales & Booth, 2007) demonstrated the sensitivity to discern among various social occasions of language use. In comparison to all other occasions for language use—from novel writing to blogging—speakers typically use words with six or more characters 9.43 percent of a transcription of a single speaking occasion. That compares to a grand mean (average average) of all language use occasions of 16.1. Thus, with 3.7*SD* (or the sample defined within broad performance bands called *standard deviations*) speakers compared to writers are -1.77 *SD* different. Such differences between speakers and writers suggested that there is far more similarity among those writing for any social occasion than there is among speakers. By consequence, the more robust dataset must derive from texts written to be read rather than primarily to be heard. Greater complexity can be expected because the reader can return to the message of the writer as often as is necessary. Thus, the writer can dare more complex statements than speakers typically make.

This greater complexity, as is argued below, casts *academic writing* as acknowledging multiple perspectives in a spirit of logical truth seeking (College Board, 2022; Flowers & Hayes, 1980; Pfister & Petrik, 1980).

2.1.2 Writing as Another Indicator of Individual Learner Differences

Differential psychologists generally support the Cattell-Horn-Carroll (CHC) model of mental abilities (McGrew, 2005). In doing so, they acknowledge two sorts of intelligence. The first comes in recognizing that solving for 7 + x = 9 suggest x can only be one number, because of the equal sign, and that it must be an integer. Accordingly, these conclusions suggest a general ability that Spearman observed among the highly positive correlations of achievement tests. Such similarity Spearman identified as *g*, while in his legacy the same general ability has also been identified as *Gf*, or a fluid intelligence associated with reasoning and problem-solving. Typically, however, children would first learn that 7 + 2 = 9 in rote memorization tasks. Assigning tables of values to memory is the sort of domain-factual intelligence known as *Gc*, or that revealed, for example, in vocabulary tests (Linn & Miller, 2005).

Accordingly, a far greater sensitivity to cognitive complexity may be approached in such thematic textual analysis programs than can be achieved in either holistically scored writing or through standardized multiple choice testing. Particularly germane to the present study were those cognitive ability items on writing achievement. The CHC model is foundational (McGrew, 2005) to commonly accepted indicators of intelligence—e.g., the Woodcock Johnson tests of Cognitive Abilities (Woodcock, Mather, & McGrew, 2001). The moderate to strong correlations among writing and reading complexity indicators for 20- to 39-year-olds on the verbal comprehension sub-test of Woodcock Johnson ranged from .5 on the incomplete words sub-test to .76 on reading vocabulary, for example. Moreover these scores

provided diagnostic commentary in the comparison of one student's test scores to others taking the same test. Thus, scores on traditional achievement instruments, then, comment on one learner's current abilities at a global level. By extension, achievement battery scores are also more robust than the variability among multiple writing samples gathered from even the same student and holistically scored.

2.1.3 The Problems with Holistic Assessment of Writing

Holistic assessment—or scoring based on a cluster of possible concerns that should weigh on a given writing occasion—provide neither the validity of traditional psychometric measurements nor assurance that all student writing samples will be graded identically. As poor as reliability—i.e., consistency—is within individual assessors, it has been found to be lower in inter-rater reliability (Linn & Miller, 2005). While the holistic assessor may be interested in how well a student appropriates the conventions of standard written English, a score could also be influenced by novelty of the audience appeal, the breadth of research in support of a point, the quality of narrative sequence, concreteness of language or any other instructor-level concern. Moreover, unless all of those components are measured in equal weight from one student paper to the next written by the same student, then the scores are not comparable.

Current empirical research underscored these points. Johnson, Penny and Belita (2000) surveyed typical score resolution procedures for achieving inter-rater reliability. These researchers found scores clustering around the middle of a 4-point qualitative scale used in scoring degrees of proficiency in writing domains. The result inflated inter-rater agreement. These same researchers also found that if adjacent scores were averaged—say an overall 2 and 3 scored on a 4-point scale—then domain-level reliability was also inflated. While inter-rater reliability procedures can be improved, Linn and Miller (2005) observed a typical problem in improving

assessment: “most of the suggestions for improving the scoring of responses to essay questions require more time, not less, as might be hoped” (p. 234).

2.2 The Social Nature of Writing

The present study operationalized writing as a social phenomenon. That possibly bidirectional relationship between language and culture noted by composition-rhetoric theory (College Board, 2022; Flowers & Hayes, 1980; Pfister & Petrik, 1980) finds complement in textual processing research. Pennabaker (2007b), for example, demonstrated that individuals who employ one of the 184 words comprising an *anger* sub-dictionary will use another anger word between 55 and 92 percent of the time within the same text. The cultural investigation complement of the software, then, in this case could find expression in the sort of language used within a broken family vs. one functionally intact. Writing, then, was seen to involve both cognitive and personality dimensions that must be managed toward doing the complex task of communicating particular content to a particular audience located in a particular place and time. I expected that those who scored high on N (e.g., alcoholics and compulsive gamblers) to write differently than those who scored high on E (those extraverted, generally enthusiastic personalities). Moreover, I held language use must also be formed by home life and the various language use occasions that frame previous learning. Sex expectations when present culturally must also influence language use. These individual differences must then partially explain differences observed in cognitive complexity as measured in writing. In what follows, I detail: (a) how *academic writing* should be understood and so guide inquiry; (b) how language use acknowledges social context and, subsequently, may influence that through the lexical (word choice) and semantic (meaning) aspects of language; and (c) how individual differences in word use can provide the basis by which to infer various levels of cognitive complexity.

2.2.1 Academic Writing Operationalized

Since audience specificity shapes the language of the message, I emphasized that cognitive complexity must be formed in consideration of audience expectations. That is as basic as to claim that those who read a poem will have different expectations than those who read a news report. The present study defined audience as both generalist and academic. By *generalist* I meant writing on topics of likely interest directed to particular stakeholders in any given discussion. By *academic* I suggested an open invitation for various perspectives to contribute to a complex discussion. I further stipulated that discussion of either *academic writing* or *academic discourse community* must be defined for the population sample under investigation, for some have argued (Caffarella & Barnett, 2000; Lewis, 2010) that academic writing is discipline-dependent, or that a particular academic audience within the Academy should guide writing to that sub-population. In the more general sense, the present study emphasized academic writing as a recursive process pulling upon both writing and reading. Badley (2009) noted academic writing is "constructing, deconstructing and reconstructing knowledge, connecting, disconnecting and reconnecting concepts, describing and re-describing our views of the world, as well as shaping, misshaping and reshaping ideas" (p. 209). In sum, then, academic writing is a recursive process embedded within a social and historical context. As such, I expected researchers can measure cognitive and affective tendencies within well-established psychometric instruments once the outcomes expected by an academic discourse community are identified.

Academic writers must further be understood as those who identify a discourse community of readers who have an inherent interest in the topic under examination. Only this degree of operationalization admitted affect into the complex transfer of meaning from writer to reader. That complexity was suggested if we grant that the effective writer must to some extent have emotional self-awareness,

even as he or she must self-regulate that writerly temperament. Because the thinker wishes to reach that audience by writing, the effective writer must consider how words can trigger either a warm or cold emotional reception.

Toward treating writing within an academic context and enfolding both cognitive and affective demands, the present research was guided by the path models depicted in Figures 3.1 and 3.2. Such models of academic writing combined the distinct languages of cognition and affect, even as it allowed for previous learning to influence language use. Among the previous learning factors are the skills already acquired as revealed by standardized tests. Also important was whatever personality temperament the learner brings forward, as was the sex effect on language use. I further expected that enrollment status must be important. In the present study that was captured by students either enrolled in a general or vocational course of study. As I expected that cognitive complexity can develop through college training, another predictor of complexity for the present study was students' matriculation status.

2.2.2 Support for Cognitive Complexity from National Standards

The present study drew upon standards jointly published by the National Council of Teachers of English and the International Reading Association (1996). The importance of writing to a particular audience (Standard Four) and to do so with multicultural tolerance for diversity (Standards Nine and Ten) supported the three-way process model of language use (Pfister & Petrik, 1980). That also complemented how I defined cognitive complexity. Moreover, the very act of writing should be taught within a communal sensibility, since the act requires acknowledgment of communal values and practices (Standard 11).

2.2.3 Cognitive Complexity

By the term *cognitive complexity*, writing researchers generally agreed that several psychological processes (e.g., information retrieval, schematization and analogy-making) evidence themselves in people's writing products, and the level in which they are evident differ greatly among writers. Delia(1982) attributed the term to Kelly (1955), who discussed cognitive complexity in the context of personal construct theory, or the theory that people mold themselves to fit various environments. Crucial to the development of such personae, Kelly argued, is postulating behavior in a socially and cognitively complex language user expects to be rewarded (Bieri et al., 1966). Complexity viewed via the Knowledge items cluster of the Woodcock Johnson battery suggests a developmental acquisition (Woodcock et al., 2001). Moreover, those effects increased with age. Processing speed was also shown to have consistent and significant effects on writing across the lifespan. In this sense, cognitive complexity was framed generally by deciphering, meaning-making and expression behaviors. Such measures, then, provided stable and global indicators of the cognitive abilities believed to indicate cognitive complexity. That is to claim that such measures of cognitive complexity demonstrated the sort of general ability Spearman (1904) identified.

At a more local level (i.e., the level of a student writing sample) researchers and theorists generally identified cognitive complexity by the use of textual markers: words, clauses and punctuation (Hunt, 1964; Pennebaker et al., 2007b; Sanders & Schilperoord, 2006). Moreover, all of the following approaches to discerning complexity in writing depended on hierarchical awareness on the writer's part. Sanders and Schilperoord (2006) argued, for example, that "a cognitively interpretable text analysis should focus on text structure rather than, for instance, on stylistic or syntactic characteristics" (p. 387) This observation suggested that cognitive complexity realized in writing is hierarchical, or that one clause can only exist in some

relation—amplification or reversal—to another. More complex writing, then, would be more hierarchically (conceptually) complex. Moreover, since complexity related to structural tendencies larger than sentences and paragraphs, then it may be considered as the outcome of what Vygotsky (1934/1986) called the higher-order psychological processes in abstraction and logic. It may also map onto what Spearman discovered, a general intellectual ability predicting similar scores by the same student on any battery of achievement tests.

2.3 Cognitive Complexity Within the Synthesis of Two Research Traditions

The Spearman-Vygotsky connection is best discussed within writing as a complex, individually realized phenomenon. As Linn and Miller (2005) observed, the only reason to use subjective assessment (essay questions) is to assess a skill a traditional psychometric instrument cannot. Yet, because Spearman expected *g* to influence performance on any human ability, we should expect *writerly g* to also mark levels of cognitive complexity.

2.3.1 Cognitive Development Emphases

Writing understood as another way of detecting individual differences joined Vygotsky and Spearman, each unknown to the other. They were in pursuit of identifying an individual, yet population discriminating variable that could account for variance within human performance. Both, moreover, were *developmental* in understanding what Spearman called the "fundaments" (1927; p. 411) of thinking and Vygotsky defined as the process of formal education. Here, the quantitative analyst who invented modern factor analysis was seen to meet the figure commonly regarded as one of the first great social constructivists. It happens they identified different aspects of the same construct: intelligence that can both be acquired and that which is apparent in pure reasoning tasks. Vygotsky prompted existing schemata through showing subjects a color card associated with words to avoid in

one of his association tasks (Dixon, 2003). Only the youngest participants benefitted by the prompt, for the older ones had a greater capacity to focus on the singular prohibition of the tasks: which words to avoid in prompting another to name an object. Spearman, by complement, knew that test scores loaded most heavily on a single factor accounting for about 75 percent of the variance among scores. The Spearman insight made possible the Cattell-Horn-Carroll model of *Gf* and *Gc*, or that denoting fluid (a.k.a., general) intelligence and crystallized intelligence or what also came to be called *procedural knowledge* (Carroll, 1993). In short, Spearman knew he had accounted for the lion's share of variance among scores on achievement tests. Vygotsky knew the end of formal education: logic and abstraction. They simply approached two questions differently, although each recognized both development and hierarchical awareness commonly frame cognitive complexity.

2.3.2 Complexity in Recognizing Audience

Vygotsky's (1934/1986) cultural-historical model of influence on language development also commented on the language user's intentional interaction with others. Complexity in this sense would be realized in making stronger, more audience sensitive arguments. Burleson and Caplan (1988) defined cognitive complexity as "an individual-difference variable associated with a broad range of communication skills and related abilities" (p. 233). Their definition seconded cognitive complexity as revealed in writing as an important marker of individual difference. Their definition also represented cognitive complexity in a way Spearman (1927) and since others (Carroll, 1993; Gustafsson, 2001; Jensen, 1992) have as a hierarchical phenomenon pulling upon skills as different as organization of new material and adapting schemata to account for differences in previously held knowledge. Finally, this complex understanding of individual language use suggested that academic writing

is realized at the intersection of present learning and appealing to an audience that values, for example, openness to examining evidence in an alternative method.

2.3.3 Other Influences on Cognitive Complexity

I held personality must also inform cognitive complexity in that occasion when the writer must not only anticipate emotional reception of a message, but also regulate his or her own emotional inclinations. Furthermore, I expected sex, enrollment status and matriculation status would influence academic writing. Accordingly, Carroll's (1993) 70-year summation of factor analysis of human abilities served the present study in the way that Cattell's exhaustive list of personality traits served as a model for possible identifiers of what became personality factors. The connection was this: the current research looked to operationalize comparison of textual processing output to the evidence-based findings apparent in standardized measures of achievement. These first-order factors indicating verbal ability tend to highly and positively correlate under a third-order factor, *g*. Second-order factors are formed in covariance among *Gf*, or fluid reasoning, factors and those relating to *Gc*, or the crystallized, vocabulary- and arithmetic-level knowledge. Thus, a review of the first- and second-order factors relating to written language and production skills follows.

2.3.4 The Hierarchical Relation of Writing and Reading Skills

Second order factors like *Gf* and *Gc* Carroll found to be influenced by LD, or language development (1993). Yet, Carroll noted, LD "is dominated by a second-order or even a third-order factor, usually interpreted as a general intelligence factor" (p. 151). By *dominated* Carroll meant that the variance uniquely associated with LD is subsumed by a higher-order factor, like *g*. To amplify the complexity of writing that Carroll found, he discussed the Verbal or Printed Language Comprehension (V) and LD as factors never entirely distinguished in his analyses. That

was to claim that the act of reading strongly covaries with writing that they might be indistinguishable. Carroll failed to find a study that distinguished Writing Ability (WA) factors from other language use abilities. Carroll wondered if LD, V and Verbal Lexical (VL) were factors that would predict the linear development of a WA factor. Toward understanding writing as a general ability, however the result of other factors, Carroll was clear: writing is "a nebulous variable that cannot be tied to any particular view of writing behavior and its antecedents and consequences" (p. 188). In this sense, WA factors form their own second-order presence in human abilities. Writing assessment, like reading assessment, must recognize the contextualization of previous knowledge and appeal to a present audience. It must be further formed to a message occasion, whether that to be convey information (exposition), argue the merit of a perspective or express some idiosyncratic message (e.g., poetry).

2.4 Measuring Cognitive Complexity in Student Writing

Generally, there have been three approaches to using written texts to measure cognitive complexity: (a) Holistic assessment of a text at some stage of development, (b) latent semantic analysis (LSA); and (c) thematic analysis. For the purposes of the present study, I employed neither holistic assessment nor LSA.

2.4.1 Holistic Assessment

Holistic assessment, or assessing by a cluster of concerns as various as adherence to conventions and exploration of various critical perspectives, was found to be so "sufficiently multilayered" (Pennebaker & Stone, 2003, p. 549) and audience driven (Emig, 1971; Flowers & Hayes, 1980) that it must be decoded by human judges. Accordingly, the problem of rater reliability came to the fore when measuring complexity qualitatively. Cherry and Meyer (1993) observed that reliability is complicated not only by the nature of holistic assessment, but also by instrument

reliability. Undoubtedly it is reasonable to expect the same student on different days to respond differently to any given prompt. While some research challenged that expectation with natural language use consistency ratings as high as .6 (Mehl & Pennebaker, 2003), the point remained that as poor as inter-rater reliability can be, then that may set the ceiling if individual learner response so radically oscillates as Spearman(1927) expected.

2.4.2 Latent Semantic Analysis (LSA)

LSA, by contrast to holistic assessment, offers quantitative commentary on complexity, but complexity understood primarily as coherence (Foltz, Kintsch, & Landauer, 2009; Landauer, Foltz, & Latham, 1998). That sense of cognitive complexity is useful, for example, if one wishes to examine how different is the language of couples who relate well in comparison to those who do not. However, the approach fails to support a learner-centered writing that licenses writers to explore topics of personal interest. By consequence, then, only thematic analysis was appropriate for the present study.

2.4.3 Thematic Analysis Approaches

Thematic analysis is a bottom-up approach (Foltz et al., 2009) to assessing text, whereby the evaluator counts the number of times a linguistic feature appears in a text. For example, a *judge-based thematic content analysis* applies a previously established coding system to find the number of times a writer discusses, for example, motive(Atkinson & McClelland, 1948), explanatory style (Seligman, Peterson, Schulman, & Castellon, 1992), conceptual complexity (Suedfeld, Tetlock, & Streufert, 1992), psychiatric syndromes (Gottschalk, Gleser, & Hambridge, 1957), goals (Stein, Folkman, & Richards, 1997), and stress reduction (Pennebaker, 1993). A variant is *word pattern analysis* realized, for example, within the Linguistic Inquiry Word Count (LIWC; Pennebaker et al., 2007b) software. This auto-

mated text analysis program provided such analysis, but without the reliability problems inherent to establishing consensus among human judges. Any new text scanned is compared to tens of thousands of texts previously analyzed on 72 dimensions of language use. So broad was the scaling process for *LIWC* that, on average, any new text's word employment will find a match 86 percent of the time within the preloaded dictionaries.

2.4.3.1 *Employing Linguistic Inquiry Word Count.* While early bottom-up textual analysis approaches (Hart, 2001; Weintraub, 1989) primarily analyzed the affective nature of language, LIWC (Pennebaker et al., 2007) allows the analyst to also explore cognitive dimensions of a text. It emerged from the search to identify how writing about, and during, a crisis can predict later health outcomes. The LIWC program references a 4,500 word dictionary and outputs 72 textual dimensions, including, for example, the number of articles used. After eliminating core words (e.g., articles and prepositions), the program calculates what percentage of the text a dimension comprises. In resonance with earlier word-count approaches, LIWC reports on large linguistic tendencies within sub-dictionaries defined by social, affective, cognitive, perceptual and biological associations. LIWC offers a functional, or pragmatic, feedback on what cognitive and affective traits are present within any digital text. The program was initially scaled by having subjects randomly assigned to write what they did in a day (highly factual writing) and those recalling an emotionally difficult experience. Such texts came from 29 studies done in 11 labs located in various English-speaking countries. On other scaling rounds, the researchers scanned 113 science articles, 714,028 blogs, 209 novels and 2,014 transcriptions of verbal speech.

2.5 Individual Differences Among Learners

Research on individual differences/differential psychology often spans three areas: (a) the development and use of measures of psychological constructs (i.e., psychometrics); (b) the structure of cognitive ability, personality, and other areas where there is variability among individuals ; and (c) and the relationship of these differences to positive life outcomes (Anastasi, 1981; Deary, 2001; Macaskill, Maltby & Day, 2002).

2.5.1 Cognitive Ability

Foundational to the study of cognitive ability was Spearman's (1904) discovery of the general factor of intelligence. To Spearman, and since to others (Burt, 1949; Deary, 2001; Eysenck, 1979) the presence of *g* means that cognitive ability has a higher-order structure, that is both *general* and what Burt and his generation offered as *special aptitudes* (i.e., dispositions meeting environmental opportunities). These researchers consistently found *g* accounting for half the differences among learners (or half the variance). Cattell and Horn (1941; 1965; 1966) expanded this discussion of *g* by identifying it is as that ability to think and act quickly, problem solve and effect short-term memory structures. They identified that capacity as *Gf*, or *g* understood as a fluid-thinking phenomenon. In contrast, Cattell and Horn stipulated the presence of *Gc*, or a sort of crystallized intelligence accompanying procedural tasks. That they identified in knowledge tests, vocabulary and acquired skills. Cognitive ability also was also found to predict academic and workplace performance, as well as social behaviors (Gottfredson, 1997, 2002) and positive health outcomes (Deary & Batty, 2006). Spearman's work led him to posit laws of individual differences (1927). He first accounted for individual differences in learner-centered dispositions. For those with a disposition to learn any content, that is likely to be triggered again. By contrast, Spearman's law of inertia stated

that cognitive processes begin and cease more gradually than their causes. The lag he identified Piaget discussed as the necessary time to assimilate new information, accommodate that to existing schemata and then come to equilibrium (Atherton, 2011). Finally, the inverse of the law of retentivity (or having a tendency to retain a disposition to learn) is the law of fatigue, or of having temporarily exhausted cognition. Spearman believed fatigue experienced in one domain could translate to others—thus the consistently positive and moderate correlations he observed among test batteries. The shift he saw between retentivity and fatigue he called the law of oscillation.

2.5.2 Personality

Personality has also been explored as a way of accounting for individual differences. The Five Factor Model (FFM) is the culmination of various research agendas effected by Guilford and Zimmerman (1976), Cattell (1946) and Eysenck(1976). As described by McCrae and John (1992), it holds there are five essential groups of personality traits under which particular traits tend to cluster. These are Openness (intellectual curiosity), Conscientiousness (organization), Extroversion (degree of social mindfulness), Agreeableness (cooperation) and Neuroticism (degree of impulse control and anxiety). Support for these personality traits is found in both natural language use and evidence-based questionnaires. Many researchers seconded Norman's (1963) adjectival taxonomy for personality as sufficient, with which McCrae and John argued he formally launched FFM as a generally accepted approach to framing individual learner differences. In the questionnaire tradition, Eysenck (1976) is credited for identifying Extraversion and Neuroticism as dimensions revealed in psychological tests. Costa, McCrae and Dye(1991) built upon that work in identifying scales constructed for Agreeableness and Conscientiousness. FFM is less a theory of personality and more of a description of hierarchical structure of large

trait groupings as revealed in factor analysis. As McCrae and Costa (1986) noted, FFM only represents the individual at a global level, or in relation to tendencies observed in many others.

2.5.3 Individual Differences as Revealed in Word Use

One marker of individual differences is the stability of a trait over time (1992). One stable trait Jensen identified is the level of g present in the individual. The likelihood of a learner scoring at a similar level, and doing so in a consistent way among tests, is so large Jensen declared, "No other kind of information concerning children's background is as highly predictive—not the socioeconomic status of the children's parents, or the parents' education, or occupation, or race, or the national origin of children's ancestry, or their gender" (p. 62). The present empirical record suggested that natural language use is another, albeit less robust, marker of individual differences and that it is also stable over time (Gleser, Gottschalk, & Watkins, 1959; Mehl & Pennebaker, 2003; Schnurr, Rosenberg, Oxman, & Tucker, 1986). Moreover, researchers also commonly acknowledged that audience, or speaking occasion, consistently shapes individual language use.

2.5.4 Personality Dimensions and Word Use

The idea that personality is related to verbal behavior clinicians have explored for decades (Furnham, Monsen, & Ahmetoglu, 2009; Sanford, 1948; Thakerar, Giles, & Cheshire, 1982; Weintraub, 1989). The empirical argument, however, remains brief. Pennebaker and King (1999) found moderate correlations ranging from .10 to .16 on each of the FFM dimensions by word choice. In general, Neuroticism was positively and moderately correlated with negative emotion word use and negatively correlated with use of positive emotion words. Extraversion had a similar low positive correlation with positive emotion words, while Agreeableness

moderately mapped onto the use of positive emotion words. Moreover, Neuroticism was predicted by higher use of first-person pronouns.

The finding that use of first-person pronoun use moderately marks Neuroticism was consistent with research that suggested *self-absorption* is another way of identifying neurotics (Davis & Brock, 1975; Ickes, Redihead, & Patterson, 1986; Stirman & Pennebaker, 2001; Weintraub, 1989). Ingram (1990) and Nolen-Hoeksema (1987) identified the robust positive correlation of depression to *self-focused attention*. That finding also called attention to the lack of a common language in personality and language use, for the FFM research followed and overlapped research done by personality and counseling psychologists. For example, Ingram reported the heritage of studying patients' *psychoneurotic* self-focus (Gottschalk et al., 1957; Lorenz & Cobb, 1953; Weintraub, 1989). Weintraub identified patients as *impulsive, compulsive, delusional, depressed, binge-eating* and *alcoholic*, but did not use the term *neurotic*. However, Pennebaker and King (1999) cited Weintraub in their FFM study of word use.

More recently, Trapnell and Campbell (1999) employed FFM to disentangle the seemingly contradictory findings that those who exhibited high personal self-consciousness also tended to display higher levels of psychological distress. In particular, positive affectivity is closely related to Extraversion (Campbell-Sills, Cohan, & Stein, 2006). Private self-consciousness, however, also similarly related with both Neuroticism and Openness(McCrae, 1993). Most recently, Rademaker, van Zuiden, Vermetten and Geuze (2011) identified *high negative affectivity* as Neuroticism.
In one of the few studies capturing the rich linguistic context of natural language use, Mehl and Pennebaker (2003) found Extroversion was the personality dimension most frequently expressed in daily living activities. Extroverts engaged in more conversations than introverts and were alone less. They also uttered more words than introverts. Agreeableness was most strongly identified in subjects' language use. In

particular, Agreeableness negatively correlated with use of swear words and positively correlated with use of first-person pronouns. The researchers also found sex effects. For example, argument evinced a moderate negative correlation on Agreeableness (-0.32) for women, while for men there was a smaller positive correlation (0.19). Conscientiousness, by complement, was most evident in subjects' daily activities and language use. Accordingly, it was positively related to amount of time spent in class, for example. Toward exploring how natural language use suggested personality differences, Yi-Tai, Chung and Pennebaker (2011) found that the first-person perspective resulted in an LIWC analysis that registered higher on emotion words. This finding was true for both writing samples done in perspective taking (assuming the voice of a first-person narrator) and in perspective switching (when writers switched to the first-person perspective).

2.5.5 Word Use Differentiated by Development

Maturity is generally associated with the development one expects to accompany age gains, even while it is also understood to express differences in experience and training. Both senses are captured when Sanders and Schilperood (2006) found mature writers exhibit a sense a greater sense of coherence in their writing than did immature writers. In writing personal descriptions, for example, mature writers introduced a topic, characterized it and then concluded. Mature writers also elaborated and evaluated further than did immature writers. Sanders and Schilperood defined maturity in differences observed among 10-, 12- and 15-year-old writers. A meta-analysis—or study examining effect and sample sizes of all relevant studies on a particular topic—conducted by Pennebaker and Stone (2003) illustrated the effect of maturation on cognition. Over 3,000 research subjects from 45 different studies representing 21 laboratories in three English-speaking nations contributed to the sample. The researchers found positive linear associations (r ranging from 0.07 to

0.26) between use of positive emotion words, future tense, words over six characters, and words connoting cognition and the increasing age of subjects. In the same two-study paper, the researchers examined the lifework of celebrated literary figures(Pennebaker & Stone, 2003). They found positive age-coefficient correlations ranging from 0.03 to 0.60 for eight of the 10 canonical writers selected. Each study suggests that development predicts cognitive complexity. In sum, the researchers found that individuals from 8 year-olds to 70-plus-year-olds came to use more positive emotion words, employed fewer negative emotion words, favored future tense and employed increasingly less past tense. By complement, most of the celebrated writers showed these same tendencies over their publishing careers. By the very act of reframing experience maturing writers evince self-regulation. Kitchener (1980) believed learning is a process of reframing cognitive dissonance, or experience that does not align with learner expectations. Such synthesis can only take place over time, and so older learners are more likely to employ the language of cognition rather than affect with distance from a dissonant event.

2.5.6 Word Use Differentiated by Sex

The influence of sex on language use has received a great deal of attention in the popular press. In the scientific community, Lakoff (1975) found sex to align with strategies in gaining power. Accordingly, she found women employing less assertive speech than men, fewer swear words, more tag phrases (e.g., "isn't it?"), more intensifiers, and more qualifications. Men were found to be more directive, precise and less emotional in their speech than women. Mehl and Pennebaker (2003) supported these findings after sampling the natural language use of 52 college students. Again, men use four times the swear words of women, more 6-character-plus words, more words associated with anger and articles. Women showed themselves to be more discrepant in using words like *would* and *could*.

Research on the influence of sex on word use spanned age groups. Heugten and Shi (2009) found that by 25 months French toddlers better processed nouns if the correct gender-marking article (*la* or *le*) were paired with a picture. Lew-Williams and Fernald (2007) report that children up to 42 months more readily identified pictures when the articles helped the children distinguish among masculine and female sequences. There was no effect for a series of pictures all calling on the same gender-marking article. Looking at the vocabularies of 14-, 24- and 36-month olds, Vallotton and Ayoub (2011) found vocabulary better predicted self-regulation than talkativeness and that girls scored higher than boys. Arthur, Johnson and Young (2007) found women used more emotionally-connotative color words than men. Brajer and Gill (2010) also found women more loquacious than men in email communication, especially when female students communicated with female professors. Felmlee (1999) observed men using half the words in evaluating a man as a woman. In a study of language use at a clinic devoted to treating depression, Fast and Funder (2010) identified women using far more self-reference words (I, me, mine) than men, $t(181) = 2.58$, $p = .01$,$r = .19$. However, sex and sexual preference did not distinguish personal pronoun usage or bodily references in a study of match-finding advertisements published by heterosexuals, gays and lesbians (Pennebaker, Groom, Loew, & Dabbs, 2004).

2.5.7 Summary

The preceding review has defined academic writing in the most generalist context of the college classroom in which the writer is expected to understand and evaluate ideas within a particular academic audience. Such audience sensitivity only becomes more demanding in domain-centered writing. Even in the case of a writer observing style conventions unique to a given discipline, use of the term *academic writing* refers in the present study to understanding language as a social phe-

nomenon. The writer enters into relation with others through common languages associated with argument, exposition and narrative. Further, the writer chooses words based on the audience to be engaged. Moreover, that diction will be mediated by the writer's understanding and acknowledgement of audience, even to the point of appreciating an audience's historic sensibility.

2.5.8 Purpose of the Current Project

The present study offered one measure of cognitive complexity to track possible differences in development of complexity over time. Toward understanding how individual differences in cognitive ability and personality may affect cognitive complexity, I posed the following questions:

1. Is cognitive complexity (CC) in an individual's writing related to performance on standardized written language placement skills tests among post-secondary students?

2. Is post-secondary matriculation, as measured by degree progress, related to CC?

3. Is there a difference in CC between students enrolled in vocational curricula versus those who take a traditional core of general education classes?

4. Do males and females differ in their average level of CC?

5. Is CC related to the personality dimension known as Neuroticism (N)?

6. What is the strength of the interaction between N and sex?

In particular, I expected: a) CC will strongly correlate with writing and reading scores taken from standardized placement tests; b) students' college credit hours will correlate with their CC scores; c) vocational students will score lower on CC than their traditionally enrolled counterparts; d) basic composition students will score lower on CC than their advanced composition peers; e) females will score higher than males on CC; f) students with higher levels of Neuroticism will also

score lower on CC because of the higher negative emotion rates contribution to that measure; and g) finally, that sex moderates the relationship between CC and N.

CHAPTER THREE

Method

3.1 Participants

The population of interest is young adults enrolled in college composition courses. By 2010 the number of degree- and non-degree seeking U.S. students had risen to 19 million students (National Center for Educational Statistics, 2013). A quantitative survey of general education coursework required at U.S. colleges found only history and philosophy courses had higher registration rates than English literature and composition (Brint, Proctor, Murphy, Turk-Bicakci, & Hanneman, 2009). Moreover, while other humanities courses were being eliminated from 1975 to 2000, English composition and literature courses were being added to catalog offerings.

The observation that English occupies an increasingly important place in general studies holds three implications for the current study. First, since Brint et al. (2009) sampled public, private and religious institutions, writing skills demonstrated and developed during college writing courses display a great deal of variability. Second, critical writing and reading continue to serve as gateway skills to earning any college degree via post-secondary institutions' composition and literature requirements. Finally, English departments continue to perform a service function in facilitating the reading and writing skills valued across the disciplines. Moreover, English departments perform this service even as other humanities programs are decreasing. Concurrent with the growth of composition and English literature courses is the growth of community college enrollment. About 35 percent of U.S. higher education enrollment occurs at the more than 1,000 communities colleges (Provasnik & Planty, 2008). When architects of the 1947 Truman Commission on

Higher Education called for increased college enrollment, they turned to the nascent junior college system, then numbering about 600 schools and sometimes enrolling as few as 75 students (Gilbert & Heller, 2013). The shift from *junior* to *community*, Gilbert and Heller argue, resulted from the Truman commission report. That shift marked a new epoch in national educational policy. For the present study, that shift underscores the importance of understanding writing development within an important component of adult education.

3.1.1 Sample Size

To determine the target sample size, I did a power analysis using the *G*Power* software (Buchner, Erdfelder, Faul, & Lang, 2009). The model I used was a difference in means between two independent groups (i.e., lower- and upper-level composition students). I used an effect size of $r^2 = .43$ based on Kobrin, Deng, and Shaw's (2011) study. I further assumed a Type 1 error rate of .05, meaning if I repeated the study multiple times, five percent of the time I would conclude there was statistically significant difference between groups when one was not present. I set the Type 2 error rate to be .90, meaning 10 percent of the time I would conclude there was not a statistically significant difference between groups when one was present. These conditions indicated I needed 230 participants.

3.1.2 Sampling Location

The sample for this study came from a community college in central Texas. Almost 98 percent identified as either African-American, Hispanic or Latino, or White; two-thirds were women, and the average age of students is 26 years (McLennan Community College, 2014). The community college typically enrolls 9,500 students during an academic year, about half of which are defined as "career training," or vocational students. A little more than half report they wish to transfer credits to a bachelor's granting college upon graduation.

3.1.3 Course Programs

The community college offers two programs of coursework, which is typical of community colleges (Porchea, Allen, Robbins, & Phelps, 2010). The first is a general education associate's degree in such fields as law enforcement, accounting and engineering. Students from these programs tend to transfer their credits to a degree program at a four-year university.

A second course of study grants students vocational certificates in such fields as digital publishing and assisted living facility management. The students may also be workers who have returned to school to retrain in a marketable workplace skill. Approximately half of the community college population is comprised by such students (Davis & Brock, 1975). Students from these vocational programs tend to move immediately into a job after graduation and tend to take fewer general education courses.

3.1.4 Writing Coursework Levels

The English department at the community college offers multiple composition courses. Most students must take two core courses as part of their general education requirements: a basic and advanced course. The focus in the basic course is exposition (i.e., delivering information), while the advanced course focuses on argumentative writing. The only students exempt from these courses are those who transfer credit or who exempt themselves through Advanced Placement performance on the English language and/or literature examinations. Still, most students take the basic composition course.

3.1.5 Recruitment

Students came to the present study through enrollment in one of the two composition courses. I contacted instructors initially through an email that was disseminated from the academic division director at the community college. That

communication was approved by a dean of general studies. A copy of that email is found in Appendix 2. The only requirements for instructor participation were use of a writing prompt sometime after the start of the semester and roughly at the same time as other instructors (the section entitled *Common Writing Prompt for Essays* contains more information about the prompt). Over the course of two semesters in one academic year, instructors variously employed the common writing prompt. For example, some instructors used it as a diagnostic instrument administered early in the semester, while others used it as an in-class final examination. Since instructors agreed to grade the assignment in any way they deemed appropriate, students received no incentive for their voluntary participation. No student took more than one course in which this study was collecting participants.

3.2 Design

The present study examined the relationship between a common outcome, cognitive complexity, and both categorical and continuous predictors. The categorical variables (i.e., factors) were: (a) sex; (b) enrollment status (e.g., traditional associates or vocational certification); and (c) course level (e.g., basic or advanced). Continuous predictors were: (a) matriculation status (i.e., number of college credit hours students had at the outset of their participation); (b) standardized test scores for reading and writing; and (c) scores on a Neuroticism measure.

3.2.1 Variables Generated from Participants' Direct Response to Present Study

3.2.1.1 *Cognitive complexity.* The present study defines cognitive complexity (CC) as the language of reason in relation with that of emotion. CC was measured using the Linguistic Inquiry Word Count (LIWC; Pennebaker et al., 2007b) software. The LIWC is an automated word analysis software, which has received considerable attention in the social sciences (Graesser & McNamara, 2012). Its utility

lies in its capacity to discern emotional states registering within any text. Those affective dimensions are complemented by other sub-dictionaries detailing cognitive operations. Among those are the languages of insight, causation and inclusion. These dictionaries are important because in this bottom-up approach the individual word is the unit of analysis (Landauer et al., 1998).

The default LIWC dictionary is comprised of 4,500 common words used to convey, for example, emotion or reasoning. In addition, LIWC output includes information on grammatical features (e.g., articles and pronouns), punctuation occurrence, verb tense, and semantic constructs as different as achievement and death. It quantifies these textual phenomena by reporting the proportion each textual dimension occupies. At present, the LWIC reports on 72 different dimensions. On average, 86 percent of the words in any given analysis text will match with the default dictionary (Pennebaker et al., 2007b).

As with most textual analysis, reliability of the LIWC's scores is difficult to measure. The evidence that has been collected tends to indicate that it is consistent in marking emotion words. For example, if an anger sub-dictionary word is used, then there is between a 92 percent and 55 percent likelihood another anger word will be used in the same text (Pennebaker et al., 2007b). The range demonstrates two methods of measuring reliability. The larger estimate comes from using all other words in the default dictionary—either another word from the anger sub-dictionary is used or not. The lower estimate comes from a comparison of all the words in the written text. The larger estimate may overstate reliability, while the lower estimate could understate it (Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007a).

The definition of CC was operationalized from the LWIC using Equation (B.1).

$$CC = \frac{cogmech}{posemo + negemo}, \tag{3.1}$$

where *CC* is cognitive complexity shaped by the demands of academic writing, *cogmech* measures the proportion of words associated with deliberative thinking processes (e.g., know, assert, evidence), *posemo* measures the positive language usage, and *negemo* measures negatively emotional word usage. Equation (B.1) captures how cognitive complexity may map onto lifespan tendencies in natural language use (Pennebaker & Stone, 2003). Specifically, cognitive word use and positive word usage tends to increase from one lifespan period to another while negative emotional word use tends to decrease. These developmental relations of reason to emotion may complement an important feature in academic writing–that academic discourse can respectfully acknowledge multiple perspectives. Developmentally, then, less mature writers may be expected to use less reasoning language and more often make negative appeals than more mature writers.

Consider the hypothetical cases of Students A and B in their respective performances on *cogmech*, *posemo* and *negemo*. Student A scores

$$A = \frac{21}{3+4} = 3\,,$$

while Student B scores

$$B = \frac{15}{2+1} = 5\,.$$

It follows, then, that a student can have a lower *cogmech* score, yet register a higher CC score as a direct result of employing less emotional language. Thus, greater affect scores will always lower CC scores. Accordingly, higher affect scores may also be understood as registering writing that is more emotionally persuasive than argumentative.

3.2.1.2 *Common writing prompt for essays.* Participants used the same writing prompt for all essays used to measure CC. This presents the same writing conditions to all study participants and enables comparison of results. Moreover, using

such a common writing prompt is typical in writing assessment when the goal is to standardize to the greatest extent common expectations and grading standards.

According to The College Board (2022), a good writing prompt not only allows a broad display of writing skills, but also is direct enough so as not to create uncertainty among topics. Further, a prompt designed to measure academic writing must be written to a common set of specifications and follow the same basic format for presentation. Moreover, the common writing prompt should be answerable by the students who will respond to the prompt. The writing task used for the current study involved (a) identifying advertisers' claims, (b) assessing those claims, and (c) considering whether or not a product or service can contribute to consumers' happiness and satisfaction. Instructors gave students assignment shown in Figure 3.1.

> Think carefully about the issue presented in the following excerpt and the assignment below.
>
> Advertisements provide information about available products and services. Many people argue, however, that something else is going on: advertisements try to convince people that when they buy things, they are also buying satisfaction and happiness. Advertisements merely fool people into believing that the next "new and improved" product will make their lives better, and the result is that people are even more unhappy and dissatisfied than they would have been without the advertisements.
>
> Assignment: Do advertisements contribute to unhappiness and dissatisfaction? Plan and write an essay in which you develop your point of view on this issue. Support your position with reasoning and examples taken from your reading, studies, experience, or observations.

Figure 3.1. Writing prompt for essays.

3.2.1.3 *Neuroticism.* Neuroticism (Eysenck, 1976, 1979; McCrae & John, 1992) was measured through 10 items taken from the International Personality Item Pool (Goldberg et al., 2006). IPIP administrators report that the reliability for the 10-item N scale is .86 (Goldberg et al., 2006). This subscale was one of five student participants completed from the 50-item group surveying Five Factor Model constructs. Table 4.1 gives all the item stems. All the items had the same response scale, which consisted of four options: (1) Nothing at all like me, (2) Not like me, (3) Like me, and (4) Just like me. I used the sum of the items for the respondents' N score.

3.2.1.4 *Course level.* There were two levels of writing courses: basic composition and advanced composition. For the analysis, I dummy-coded this variable as basic (0) and advanced (1).

3.2.2 Variables Collected from Institutional Records

The following variables were collected from institutionally archived data: (a) sex, (b) standardized test scores of reading and writing, (c) enrollment status, and (d) matriculation status.

3.2.2.1 *Sex.* For the analysis, I dummy-coded the sex variable.

3.2.2.2 *Standardized reading and writing assessments.* Reading and writing skills were measured by the Accuplacer placement examination (College Board, 2003). The Accuplacer is a test battery composed of nine subtests designed to profile a student's previous learning and aid in placement decisions. Each subtest provides items capable of discriminating from high proficiency to low proficiency examinees. For this study, the subtests used were: (a) *Reading Comprehension* (RC), and (b) *Sentence Skills* (SS).

The RC subtest prompts students to identify main ideas from reading passages, discern direct from secondary ideas, make inferences, apply ideas and determine sentence relationships. The SS subtest measures how well a student recognizes what constitutes a complete sentence and details students' capacity to coordinate and subordinate sentence clauses, as well as their ability to restate sentence-level logic shifts in a longer reading passage. Combined, these subtests are designed to measure how well students have learned the conventions of standard English and general argument strategies. These scores have shown high test-retest reliability, ranging from .76 to .90 for RC and from .73 to .83 on SS (College Board, 2003). As a whole, the Accuplacer has previously demonstrated modest success in predicting how well students will do in the courses into which they are placed (Mattern & Packman, 2009). The combined RC and SS scores showed a modest correlation of .24 in predicting course success.

3.2.2.3 *Enrollment status.* Two categories of enrollment defined participants in the present study. I defined Workforce as those students taking a vocational curriculum approved by a state employment office. The second enrollment status is Transfer and was defined by students seeking associates' degrees rather than vocational certificates. It is not core content that distinguishes the degree from a certificate. A student at the community college, for example, could earn either an associate's degree or a certificate in digital publishing. However, to earn the associate's, that student would have to take additional general education courses. For the current analysis, I dummy-coded this variable.

3.2.2.4 *Matriculation status.* I measured matriculation progress by the number of college hours completed at the time of students' inclusion in the present study.

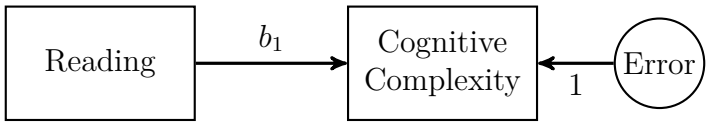
3.3 Procedure

3.3.1 Data Collection

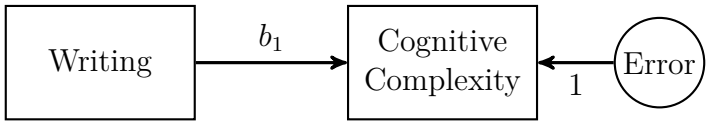
Participants had the option of completing the instruments for the project electronically or via paper-and-pencil instruments. The paper-and-pencil instruments were initially distributed to participating instructors and then picked up after students submitted their responses to the common writing prompt for assessment. Only those accompanying the voluntary consent form were retained for final analysis. Approximately 30 percent of the data were collected electronically through a secure online survey, at which students completed a consent form and personality profile as well as uploaded their essays.

3.3.2 Data Analysis

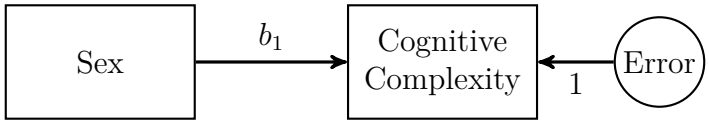
Path models for each research question are shown in Figures 3.2 and 3.3. The outcome variable for each model was cognitive complexity. I used regression for all data analyses. For the first five research questions, I used simple regression. For the sixth question, I used a multiple regression model with an interaction term.



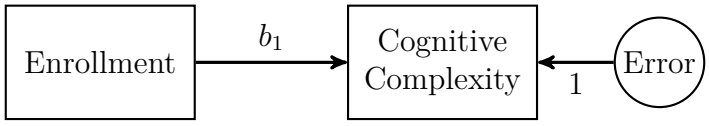
(a) *Relation between cognitive complexity (CC) and a standardized reading assessment.*



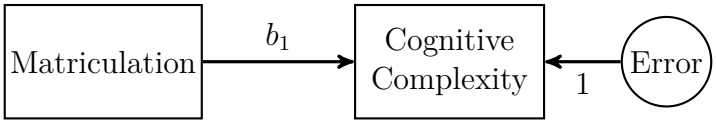
(b) *Relation between cognitive complexity and a standardized writing assessment.*



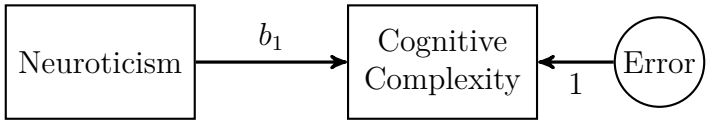
(c) *Relation between cognitive complexity and sex.*



(d) *Relation between cognitive complexity and enrollment.*



(e) *Relation between cognitive complexity and college credits.*



(f) *Relation between cognitive complexity and Neuroticism.*

Figure 3.2. Path models for research questions requiring a simple regression.

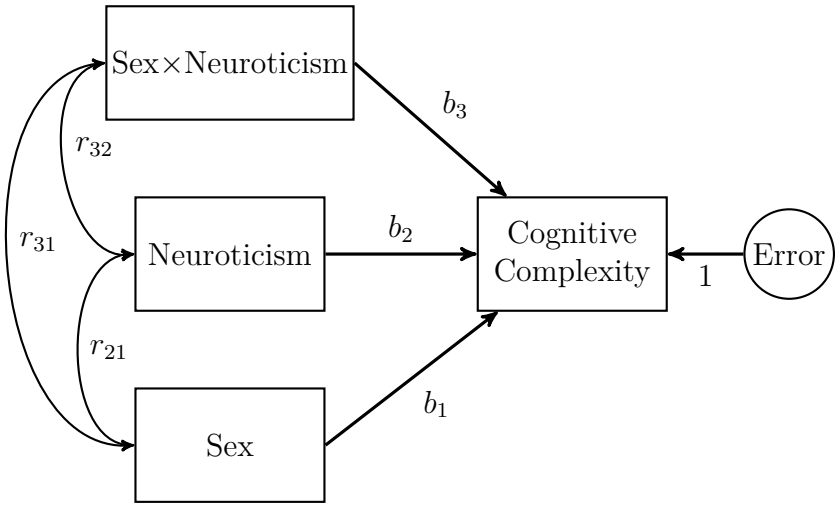


Figure 3.3: Path model of the influences of the personality construct Neuroticism, the categorical variable sex and that interaction on cognitive complexity.

3.3.3 Data Analysis Program

All analyses were completed in **R** (R Core Team, 2013). I used the following packages within **R**: (a) `BaylorEdPsych` (Beaujean, 2012), (b) `compute.es` (del Re, 2013), (c) `mice` (van Buuren & Groothuis-Oudshoorn, 2011), (d) `MissMech` (Jamshidian, Jalal, & Jansen, 2013). (e) `psych` (Revelle, 2014) and

3.3.4 Missing Data

All variables but those derived from automated text analysis had missing values (see Table 4.1). As missingness can bias estimates and thus weaken inferences that can be made, I first established the nature of missingness (Enders, 2011; Little, 1992). There are generally recognized three missing data mechanisms (i.e., how the probability of a missing value relates to the data). The first is *missing completely at random* (MCAR). Missingness in this pattern is neither related to the variable in which there are missing values nor any other variables. If missingness is in some way related to other variables collected in the dataset, then the data may be *missing at random* (MAR). Finally, if the nature of missingness is a result of the

variable itself (e.g., the under-reportage seen frequently in socio-economic status), then the pattern is *not missing at random* (NMAR).

Following best practices in handling missing data (e.g., Schlomer, Bauman, & Card, 2010), I first determined the nature of missingness in my study by examining the quantity and patterns of the missing values. That analysis revealed 15 patterns of missing data, with 1208 total missing values. I then estimate Little's (1988) χ^2, which is a statistical test that compares the mean values among each unique pattern of missingness as well as those not missing any values. If these values are radically different, and the χ^2's p-value is $< .05$, this indicates the data are likely not MCAR. In addition, I used a newly-developed test for MCAR data that examines both means and covariances between the groups (Jamshidian & Jalal, 2010). The value of Little's (1988) statistic was $\chi^2_{df=159} = 165.77, p = .34$. Moreover, the test of means and covariances was also non-significant ($p = .13$). These results indicate that the data were likely MCAR.

To handle the missing data, I employed multiple imputation (MI). That is a statistical procedure that establishes both plausible and possible values for participants' missing information. Baraldi and Enders (2010) referred to MI as a breakthrough in the last century of researchers' choices in dealing with missing data, and it is frequently cited as one of the best methods for accounting for missing values (e.g., Jelicic, Phelps, & Lerner, 2009). Enders (2011) gives a detailed account of the MI process, but the general idea is that multiple values for the missing values are generated across multiple datasets. The statistical analysis of interest is then computed in each dataset, and then the parameter estimates and their standard errors are pooled together.

Schafer (1997) recommended using three to five imputed datasets, although Schlomer et al. (2010) noted that because of rapid computer processing speeds now available to researchers "there is little drawback to selecting a larger number" (p.

5). Graham, Olchowski and Gilreath (2007) made the stronger argument that the power to make an inference decreased based on two factors: (a) the degree of missingness and (b) the number of imputations performed. They reported that for 50 percent missingness, the power to detect a false hypothesis decreased from 78 percent for 100 imputations to 59 percent for three. In the present, I used 50 imputed datasets.

3.3.5 Effect Sizes

To interpret the study's findings, I used three related measures of effect size: (a) Pearson correlation, r; (b) squared Pearson correlation, r^2; and (c) standard mean difference, *d*. I employed the r statistic for three reasons. First, as Fritz, Morris, and Richler (2012) argue, it is a familiar way to examine the relationship between two continuous variables. Furthermore, squaring that statistic yields r^2, or an indicator how much variance one variable accounts for in its relation to another.

Cohen's (1992) d serves as an indicator of the strength of association between a continuous variable and a categorical variable (i.e., factor) that has two level (e.g., male and female). As with r, d provides a standardized metric to describe the results across variables that is independent of sample size. It is calculated by dividing the between-group differences in mean values for a variable by the pooled standard deviation of the variable for the two groups. Consequently, d's metric is in standard deviation units and can register differences beyond +1 or -1.

CHAPTER FOUR

Results

4.1 Participants

There were 222 participants in this study. They were all enrolled in a Central Texas community college in the 2012-2013 academic year. Data collection occurred over both the spring and fall semesters. In Table 4.1 I summarize the descriptive statistics for variables in the present study.

4.2 Simple Regression

In Table 4.2 I present the relations between CC and the predictor variables. I list the predictor variable, unstandardized regression coefficient (B), standard error of the unstandardized regression coefficient (SE), 95% confidence intervals upper- and lower-limits, and the Pearson correlation coefficient r.

4.2.1 Standardized Reading and Writing Assessments

Reading was negatively related to CC ($r = -0.14$, 95% CI: $-0.12 - +0.15$). The two variables shared approximately 2% of their variance. Likewise, Writing was negatively related to CC ($r = -0.28$, 95% CI: $-0.05 - -0.21$). The two variables shared approximately 8% of their variance.

4.2.2 Matriculation Status

Matriculation was positively related to CC ($r = 0.05$, 95% CI: $-0.13 - +0.13$), indicating participants who earned more college credit hours had slightly greater CC scores. The two variables shared approximately 1% of their variance.

Table 4.1. Descriptive Statistics.

Variable Name	Variable Description	n	M/ count	SD/ %	% missing
Sex	Female	186	110	59	15
Enrollment	Workforce	164	19	12	26
Matriculation	Total college credit hours	89	24.50	12.37	59
Writing	Accuplacer sub-scale	88	97.75	14.82	64
Reading	Accuplacer sub-scale	89	90.50	10.24	63
Posemo	Positive emotion words	222	3.30	1.39	0
Negemo	Negative emotion words	222	2.03	1.29	0
Cogmech	Cognitive mechanics words	222	18.45	5.78	0
Stress	I stress out easily	153	2.29	0.95	31
Relaxed	I am relaxed most of the time	153	2.12	0.82	31
Worry	I worry about things	153	1.88	0.83	31
Change	My mood changes a lot	146	2.49	0.90	34
Optimistic	I often feel optimistic	153	2.10	0.69	31
Disturbed	I am easily disturbed	146	2.60	0.83	34
Upset	I get upset easily	147	2.70	0.93	34
Change	My mood changes a lot	146	2.49	0.90	34
Moodswing	I have frequent moodswings	146	3.09	0.84	34
Blue	I often feel blue	142	2.82	0.71	36
Neuroticism		145	24.51	33	34

Note. n = number of cases; M/count = mean/participants; SD/% = standard deviation/percent of reference participants.

Table 4.2. Results From Simple Regressions Predicting Cognitive Complexity.

				CI	
	B	*SE*	*r*	lower limit	upper limit
Reading	-0.01	0.01	-0.14	-0.12	0.15
Writing	-0.03	0.01	-0.28	-0.05	0.21
Matriculation	0.01	0.19	0.10	-0.13	0.13
Enrollment[a]	0.07	0.35	0.01	-0.13	0.13
Sex[b]	-0.44	0.26	-0.24	-0.01	0.25
Level[c]	-0.01	0.01	-0.14	-0.11	0.15
Neuroticism	-0.00	0.02	-0.01	-0.04	0.04

Note. B = unstandardized regression coefficient; SE = standard error of B; r = Pearson correlation; CI = confidence intervals of r. All categorical variables were dummy-coded.
[a.] Traditional curriculum coded as reference group.
[b.] Female coded as reference group.
[c.] Basic composition course coded as reference group.

4.2.3 Enrollment Status

Participants' enrollment status (i.e., traditional associate-degree-seeking or vocational) positively related with CC. As traditional associate-degree-seeking students were coded as the reference group, the positive relationship indicated that, on average, students taking vocational curricula exhibited higher CC scores. More specifically, vocational students exhibited an average CC score of 20, while their associate-degree-seeking peers average 18.2. This translates to a d value of 0.03 (95% CI:$-0.31 - +0.40$).

4.2.4 Sex

Participants' sex exhibited a negative relationship with CC. Females were coded as the reference group, so the negative relationship indicated that, on average, females demonstrated higher CC scores than males. Specifically, female college students exhibited an average CC score of 4.34 while males averaged 3.84. This difference translated to d value of 0.24 (95% CI:$-0.06 - +0.54$). While the 95% CI

for d encompassed zero, the majority of the CI was larger than zero likely indicating that females had higher levels of CC than males, but this study was not able to measure the magnitude of the difference very precisely.

4.2.5 Course Level

The relationship between course level and CC was positive. The basic composition course was coded as the reference group, so the positive relationship indicated that, on average, CC scores from students in the basic composition course were higher than those from students in the advanced composition course. Specifically, students in the basic course exhibited an average CC score of 4.21, while students in the more advanced course averaged 3.89. This difference translated to a d value of 0.18 (95% CI: $-0.10 - +0.46$). While the 95% CI for d encompassed zero, the majority of the CI was larger than zero likely indicating that students in the basic composition course had higher levels of CC than students in the advanced composition course, but this study was not able to measure the magnitude of the difference very precisely.

4.2.6 Neuroticism

Before answering whether Neuroticism was related to CC, I first examined the reliability of the N score in the current study's sample.

4.2.6.1 *Reliability of neuroticism score.* Score reliability is most frequently measured using Cronbach's (1951) α (Sijtsma, 2009). α assumes the items all measure a single construct, and do so equally well (i.e., no item is a better measure of the construct than another). This is seldom the case with items from psychological instruments. Moreover, one of the fundamental assumptions for α is that the data are continuous, yet the IPIP items use a four-point ordinal scale. Gadermann, Guhn, and Zumbo (2012) note that estimating the reliability of an ordinal scale

with α tends to bias estimates. Consequently, I used two other estimators of reliability: ordinal α and ω.

Ordinal α (Gadermann et al., 2012) estimates reliability using correlations among variables defined by ordered categorical items. Such items typically range from *strongly agree* to *strongly disagree*. That stands in contrast to the correlation matrix used to estimate reliability within Cronbach's α, which is only appropriate for continuous data. Gadermann et al. (2012) note that the use of that correlation procedure with ordinal data tends to underestimate reliability.

The second reliability coefficient I report is McDonald's (1999) ω. It is appropriate because ω is an easily calculated and understood expression of lower-bound reliability (Revelle & Zinbarg, 2009). Specifically, ω estimates the amount of variance in the items that is explained by the construct they are designed to measure. Revelle and Zinbarg (2009) found ω to well estimate the total common variance of a test in comparison with other estimators. This statement of common variance is important, they explained, because rather than being concerned only with "the greatest lower bounds as estimates of a reliability of a test, we should also be concerned with the percentage of the test that measures one construct" (p. 3).

McDonald (1999) summarized ω as the ratio between the common factor of a group of a test items and their total variance. From this definition he concluded that ω will always be a superior estimator to Cronbach's α because α is a lower bound estimate to ω unless all items measure a single factor and equally relate to that single factor. Only in that case is α equivalent to ω. α, moreover, assumes that all items have the same mean, the same item variance and the same measurement errors (Dunn, Baguley, & Brundsen, 2013). ω, by contrast, allows the means, variances and errors to vary. Its singular assumption of unidimensionality makes the estimator far more applicable than α.

In summation, both ω and ordinal α more accurately estimated reliability than Cronbach's α in the present study. Ordinal α was .85. In complement to that estimator, McDonald's ω was .84. By contrast, Cronbach's α was .63, which is substantially lower than both other reliability estimates. The ω and ordinal α reliability values in the present study exceed the .80 minimum threshold for basic research advocated by Nunnally and Bernstein (1994).

4.2.6.2 *Neuroticism and cognitive complexity relationship.* N negatively related to CC (r = -0.01, 95% CI: 0.04 − 0.04), indicating that participants who scored higher on the N test had lower CC scores. That relation, however, was the smallest negative relationship revealed in the present study. The two variables shared less than 1 percent of their variance. I next examine is sex moderated the relationship of N and CC.

Table 4.3. Results of Regression Examining Sex-Neuroticism Interaction.

				95% *CI*	
	B	*SE*	B^*	Lower Limit	Upper Limit
Intercept	3.98	0.73	–	2.54	5.42
Sex	0.16	1.08	-0.04	-1.96	2.29
Neuroticism	0.01	0.03	0.04	-0.04	0.06
Sex-Neuroticism Interaction	-0.03	0.04	-0.11	-0.11	0.06

Note. R^2 for model with only main effects is .01. R^2 for model with main effects and interaction is .02. B = unstandardized regression coefficient; SE = standard error; B^* = standardized regression coefficient; CI = confidence interval for unstandardized regression coefficient.

4.2.6.3 *Interaction of sex and neuroticism on cognitive complexity.* The results of the interaction analysis are given in Table 4.3. The R^2 value for the regression model with the interaction was twice as large as the model with the main ef-

fects alone. Nonetheless, the interaction model only explain 2% of the variance in CC. Moreover, the 95% CI for the interaction term contained zero. Thus, a very small interaction effect may be present, but the current study was not able to measure such an effect very precisely. A graphical depiction of the moderating influence is shown in Figure 4.1.

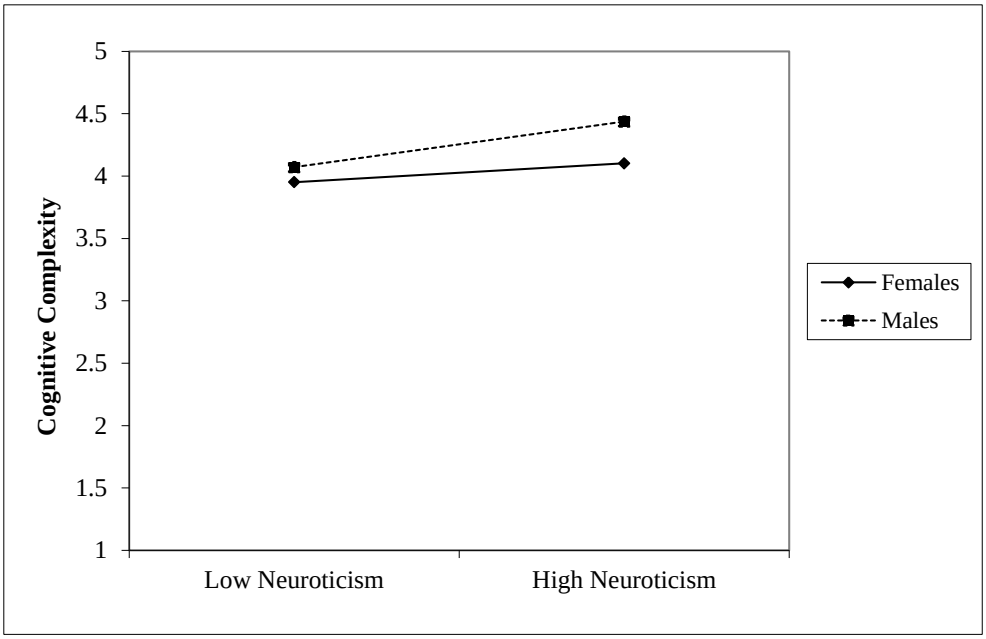


Figure 4.1: Interaction between Sex and Neuroticism. The low and high values of Neuroticism are $\pm$ 1SD from the mean.

CHAPTER FIVE

Discussion

5.1 Findings From the Current Study

In the present study I examine how a measure of cognitive complexity (CC) relates with individual learner identifiers as different sex and the personality dimension known as Neuroticism. That measure is a composite of reasoning and affect. While I hypothesize that CC would be smaller for vocational, basic composition and female students, the data do not support those expectations. In fact, I find vocational students have higher values on the CC variable ($d = 0.03$) than their associates' degree-seeking peers. Females score higher than males ($d = 0.24$), while basic composition students outscore their advanced peers ($d = 0.18$). I hypothesize that standardized testing scores on reading and writing would positively correlate with CC scores. The data show, by contrast, that there is a negative (r = -0.14) relation of reading to CC, while there is a larger but also negative relation of writing to CC (r = -0.28). I expect the personality construct identified as Neuroticism (N) to negatively correlate with CC. The data fail to support that expectation. Higher Neuroticism positively attends higher complexity scores. However, there is a negative relation between N and CC when moderated by sex (r = -0.10). I hypothesize that more exposure to college would result in higher CC scores. The data also support that expectation. Greater college earned at the outset of the present study show a positive relation to CC ($r = 0.10$).

5.2.1 Academic Writing in General Coursework

One connection of the present study to the current literature comes in the consideration of academic writers forming a distinct discourse community. That may be shown in the contrasts demonstrated in basic and advanced students, as well as between degree-seeking vs. vocational students. In each case, greater exposure to general college coursework exhibited lower CC scores. Yet, those scores may be an indicator not of advancing performance in reason and emotional self-regulation, but in the typicality of language use.

Toward exploring the possibility that CC may actually be an indicator of typicality rather than development of cognitive reasoning, I consider the LIWC norming sample for scientific papers comprised of 113 technical articles gathered from the journal *Science* from 1997 to 2007. That corpus represents a comparable discourse community. That comparison highlights tendencies within the academic discourse community of college students preparing to transfer to four-year institutions and enter (or re-enter) the work force. On the whole, advanced composition students bear more resemblance to the technical writers of *Science* than they do to their basic composition counterparts. That conclusion is drawn from two indicators of typicality, as well as one linguistic feature that may illustrate novice tendencies.

The two indicators of typicality are use of function words and the proportion of words matching the LIWC default dictionary. *Function* words (e.g., *this, that, an, with, no*, and *why*) serve a grammatical function in typical usage (Weber, 2006). That is to observe that these words function to announce nouns (e.g., *a, an, the*) or indicate nouns as objects of prepositions (e.g., *above, below, beyond*). Common function words are also pronouns, conjunctions and auxiliary verbs. They represent about .4 percent of the typical 100,000-word vocabulary the average English-

language speaker employs (Weber, 2006), yet they comprise about 50 percent of the words daily used. LIWC features 443 of these most commonplace words.

The *function* words employed in *Science* articles comprise about 35 percent of the total word count (Pennebaker et al., 2007a). By comparison, *function* comprises 38 percent of the texts advanced students compose, while it represents 57 percent in basic-level compositions. In sum, basic composition students bear far more resemblance to novelists whose work assisted in LIWC norming ($M = 57.17$) than to writers contributing to an academic discourse community.

Additional support for basic composition students establishing a baseline for typicality is seen in their *Dic* scores. The *Dic* score is the proportion of text matching to the 4,500 words comprising the default dictionary register. On average, basic composition students match 88 percent of their words with the LIWC default dictionary, while advanced students match only 62 percent and *Science* writers even fewer (54 percent).

Another LIWC variable may be understood to indicate atypicality. That variable is *Sixltr*, or words with six or more characters. Pennebaker and King (1999) found higher use of longer words moderately correlated with SAT scores. Yet, higher education culture can encourage students to use overly complex words in a bid to sound more intelligent (Oppenheimer, 2006). Basic composition students used slightly more complex words than did their advanced counterparts on average (21.12 vs. 19.78), while *Science* writers used about a third more complex words on average (29.56) than did advanced composition students. These comparisons suggest there is occasion for using atypical language (e.g., scientific writing). Yet, by the discipline development reasoning that makes atypical word use the norm for a particular field, basic composition students should have less occasion to use atypical words than their advanced composition peers.

The comparisons highlight what Biber (1992) found. Complexity may be more readily identified in the situational demands of communication (e.g., a speech in which audience members are free to interact with the speaker) than in the form of communication (exposition vs. personal narrative). Advanced college writers in the present study may sense a greater inclusion within an academic discourse community and so more readily emulate their professors. Basic college writers, by contrast, may rely more on the commonality of function words because those more greatly inform their operative daily vocabulary.

The present study's attention to sex differences maps onto other studies. Ginsburg, Ogletree, Silakowski, and Tammy (2003) documented men's greater tendency to use coarse language, while Leaper and Robnett (2011) examined 29 similar studies and found a small positive effect ($d = .26$) for tentativeness more evident in women's speech. Arthur (2007) reported women using more emotionally-connotative words, while Brajer and Gill (2010) identified women using more words in emails. Women have also been observed using more self-reference words than men (Fast & Funder, 2010). By complement, I find women employ more positive emotion words and fewer words associated with either negative emotion or reasoning. In sum, these preliminary findings may underscore how word use can exhibit power, vent tense emotionality and further reinforce gender differences.

The empirical literature on development of writing during the college years is relatively sparse and more diverse in foci. For example, recognizing plagiarism and avoiding it may be considered developmental advances (Barry, 2006; Owens & Fiona, 2013). Retraining students' attribution of success and failure may also distinguish between those who earn college degrees and those who do not, Hall et al. found (2004; 2007). Maclellan (2004) ordinally ranked degrees of critical reflectivity. One other example of establishing complexity in reasoning and writing looks at the task of synthesizing material from sources outside the writer. Mateos and

Solé (2009) report on the difficulty of the task even for college students. Younger students tended to synthesize texts sequentially, or as they read them, while older students thematically grouped texts after reading a diverse selection selected for difficulty. These various studies relate to the present study in a common effort to distinguish what marks reasoning and socio-emotional regulation for adult learners.

5.2.2 Manifest and Latent Expressions of Cognitive Complexity

The present study resonates within a growing literature of establishing degrees of cognitive complexity two different ways. Biber (1986; 1988; 1992) and Biber and Finegan (1989) examined manifest lexical features of natural language usage. They found complexity is moderated more robustly by whom is speaking than it is by speaking occasion. What he calls *discourse complexity* (Biber, 1992) is moderated by speaker-audience interaction, the purpose of speaking (i.e, relaying information or telling a story) and audience expectations in delivery (e.g., passive, third-person accounts of scientific research). A second approach to identifying complexity marks, for example, degrees of cohesiveness among texts under analysis (Landauer & Dumais, 1997; Landauer et al., 1998). More important than surface dimensions of text is how similar, for example, are medical students' discussion of heart transplants compared to textbook writers. This approach finds complexity, then, in similarity to expert language use.

In the present study I employ both manifest and latent tendencies to examining student writing. The indicators of reasoning and feeling I use to indicate cognitive complexity come from very words student writers employ. That is a manifest approach. By complement, I also explore how word use signals greater inclusion within an academic discourse community. That inclusion, at present, can only be approximated through what can be observed—indicators of emotional self-

regulation and reasoning. Thus, I also do analysis of complexity through latent constructs such as *critical thinking, academic writing* and *literacy*.

5.3 Limitations of the Present Study

I recognize several phenomena affecting the generalizability of the present study. Even though I asked community college instructors to administer the common prompt within the first half of the semester, only about half did so. The others did so at the end of the fall and spring semesters in the form of an in-class final examination.

Complicating the matter of offering a typical sample of community college writing, the present study included dual-credit high school students. That means these students received both high school and transferable college credit for completing one or both of the composition courses. These students are possibly atypical in the fact that not all planning to attend college would necessarily enroll in the local community college.

Another limitation is the narrow developmental spectrum sampled in the present study. It is conceivable that advanced composition students could be in their first semester at the community college and so show little exposure to college coursework. Such students are among others who are not only older (i.e., veterans attending college for the first time and those retraining for employment), but also those who have had much greater exposure to college coursework. Even if there were not such confounding issues, and if only those students of a similar age and coursework history were admitted to the study, there may yet be one other matter detracting from the generalizability of the present study. One to two years may not be enough time for students to assimilate and employ learning in order to demonstrate change. Finally, the word groups of negative and positive emotional expression may have different statistical distributions and so should not be considered aggregately in relation to the language of reasoning (i.e., the *cogmech*variable). Pen-

nebaker and Stone (2003) found that positive emotion and reasoning increase over a lifetime, even as negative emotional language usage decreases. Relating both positive and negative affect to reasoning, then, may conflate developmental tendencies if those are present to identify.

5.4 Directions for Future Study

In consideration of the above stated limitations, and in recognition of the developmental and individual learner differences identified, I would further investigate the union of thinking and feeling revealed in writing by:

1. Collecting at least four writing samples per participant to gather a more stable cognitive and affective profile of each participant;

2. Having collaborators effect study through more uniform conditions;

3. Sampling a broader range of age- and performance-participants;

4. Examining cognitive complexity as the ratio of cognition to positive affect separately from that of negative affect; and

5. Establishing the difference in cognitive complexity among not only writing genres (i.e., non-fiction vs. poetry), but also across academic disciplines toward gathering baselines for both rhetorical occasion and audience.

APPENDICES

APPENDIX A

R Syntax Used for Current Study

```
# missing data patterns

library(mice)

MVA <- read.table("Ch4analysis.csv",header=TRUE,sep=",")

md.pattern(MVA)


# Little's test for MCAR

library(BaylorEdPsych)

mcar.little <- LittleMCAR(MVA)


# examining mean and covariance differences among missing data patterns

library(MissMech)

TestMCARNormality(MVA)


# regression using mice package

library(mice)

sex <- with(imp, lm(CC~sex ))

enrollment <- with(imp,lm(CC~enrollment))

writing <- with(imp,lm(CC~writing))

credits <- with(imp,lm(CC~credits))

reading <- with(imp,lm(CC~reading))

level <- with(imp,lm(CC~Level))

interaction <- with(imp,lm(CC~sex + Stability + sex*Stability))


# Cohen's d with compute.es package
```

```
library(compute.es)

#factored by sex, for example#

tes(t=.75, n.1=57, n.2=32)


# multiple imputation with mice package

library(mice)

MVA #name of analysis file#

imp<-mice(MVA,m=50)

imp

complete(imp)


# reliability of IPIP items with psych package

library(psych)

alpha(examplename$rho)#examplename is a analysis file place holder#

# Reliability analysis

library(psych)

fa(reliability2)#reliability2 is name of analysis file#

guttman(reliability2)

omega(reliability2)
```

.

APPENDIX B

Initial Email Communication to Community College English Faculty

2.0.1 Measuring cognitive complexity

To get at writing as a way of understanding individual differences, I will use an automated textual analysis program (LIWC 2007). This software outputs 70 lexical and semantic features of written language. The key markers of complexity come from the software's comparison dictionaries devoted to the distinct languages of cognition, positive emotion and negative emotion. Each value defines the proportion of the text any given dimension occupies. I will define cognitive complexity as

$$CC = \frac{cognition}{posemo + negemo}, \tag{B.1}$$

where *cognition* is the default indicator of thinking, *posemo* is the indicator of positive emotion and *negemo* is the indicator of negative emotion. Expectations of tolerance for and appreciation of diverse perspectives in academic writing inform this equation. Since there is no gold standard for how much negative or positive emotion should frame academic writing, I admit the whole range of emotion as possibly influencing complexity.

2.0.2 Sampling requirements

To understand how this new measure of cognitive complexity relates to existing placement measurements, I need to sample as many different sorts of student writers as I can–from those in developmental sections to those completing their requirements for graduation. Then I can answer the question that most concerns MCC–Do online vs. face-to-face writing students differ in learning outcomes?

2.0.3 Research design

I will gather writing samples from developmental students through those completing English 1302 and sophomore literature courses. Differences among levels will be determined through a common writing prompt recently published by the College Board.

2.0.4 Instructor requirements

This prompt will be administered about halfway into the semester. Instructors will choose how much, if any, to weight it. Instructor participation simply requires that the prompt be used in some way so that college writing instruction has had some time to influence learners. The researcher requests that the prompt be graded as instructors normally would. If it is not graded, then the researcher needs to indicate that in his analysis. Instructor commitment is about 10 minutes. That gives me time to briefly recruit students and for instructors to post my link on Blackboard.

2.0.5 Student requirements

Students are only required to submit a paper for analysis and complete a personality profile required to contol for individual learner differences. All students who do so will be eligible to receive gift certificates donated by local merchants. Participant names will be randomly drawn to award incentives. I expect students will complete the online personality survey in about 15 minutes.

2.0.6 The College Board prompt

Advertisements provide information about available products and services. Many people argue, however, that something else is going on: advertisements try to convince people that when they buy things, they are also buying satisfaction and happiness. Advertisements merely fool people into believing that the next *new and*

improved product will make their lives better, and the result is that people are even more unhappy and dissatisfied than they would have been without the advertisements.

Assignment: Do advertisements contribute to unhappiness and dissatisfaction? Plan and write an essay in which you develop your point of view on this issue. Support your position with reasoning and examples taken from your reading, studies, experience, or observations.

2.0.7 After the study

After completing analysis, I will be glad to release a general report to the whole of the MCC network. I will also be available to discuss instructor-level benefit of employing automated textual analysis within assessment systems.

APPENDIX C

Informed Consent Document

C.1 The Baylor Educational Psychology Department

C.2 Primary Researcher: David Thomson

The purpose of this study is to determine if a measure of writing complexity relates to other known methods of placing students in the appropriate classes. Doctoral candidate David Thomson will also examine possible influences on writing through enrollment, graduation progress and personality.

The results of the research may be published or presented at a conference, but your name will not be used. All information obtained in this study will remain confidential to the extent allowed by law.

No foreseeable risks or discomforts are associated with this study. You should understand, however, that your English instructor by agreeing to allow me to contact you has agreed to grade a writing prompt I have provided to all participating instructors.

I will ask an authorized MCC representative to match your study identification number only with the following data:

1. how many hours you have completed at the time of your participation;
2. what you have reported as your race/ethnic identity and sex;
3. whether or not you are among the students who take one of the Work Force courses of study;
4. what are your reading and writing scores on either the Accuplacer or another standardized and comparable test that could be used for college placement;
5. your date of birth.

Your data will be kept under lock and key after it has been collected. Also, your data is confidential as allowed by law. Your name will be used only to record your consent and participation. All confidential data will be kept in a locked office in the Department of Educational Psychology at Baylor University, and will be destroyed after the results are published.

As you may be aware, electronic communication may be subject to interception in transit. Therefore, it is possible that your information might be seen by another party. I cannot control whether or not that happens. If you are concerned about your data security, I suggest that you record your answers to the survey in the following manner: Number [space] letter or number, essay or yes-no-response at david_thomson@baylor.edu.

If you have any questions about your rights as a participant, or any other aspect of the research as it relates to your participation, please contact Dr. David W. Schlueter, Ph.D., Chair Baylor IRB, Baylor University, One Bear Place 97368 Waco, TX 76798-7368. Dr. Schlueter may also be reached at (254) 710-6920 or (254) 710-3708.

"I give my consent to participate in the above study. Furthermore, I understand that I must be 18 years of age or older to be included in the study. Moreover, I understand that minors will be removed from the study. Finally, I understand that as a voluntary participant, I can withdraw from the study without penalty or loss of benefits."

RETURN TO RESEARCHER

REFERENCES

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological monographs*, *47*, i-171.

Anastasi, A. (1981). *Differential psychology* (4th ed.). New York, NY: Macmillan.

Arthur, H., Johnson, G., & Young, A. (2007). Gender differences and color: Content and emotion of written descriptions. *Social Behavior & Personality: An International Journal*, *35*, 827-833.

Atherton, J. S. (2011, 12 November 2011). *Learning and teaching; Piaget's developmental theory.* Message posted to http://www.learningandteaching.info/learning/piaget.htm.

Atkinson, J. W., & McClelland, D. C. (1948). The effect of different intensities of the hunger drive on thematic apperception. *Journal of Experimental Psychology*, *38*, 643-658.

Badley, G. (2009). Academic writing as shaping and re-shaping. *Teaching in Higher Education*, *14*, 209-219.

Baraldi, A. N., & Enders, G. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, *48*, 5-37.

Barry, E. S. (2006). Can paraphrasing practice help students define plagiarism? *College Student Journal*, *40*, 377-384.

Barthes, R. (1972). *Mythologies*. New York, NY: Hill and Wang. (Original work published 1959)

Beaujean, A. (2012). *Bayloredpsych: R package for baylor university educational psychology quantitative courses.* http://CRAN.R-project.org/package=BaylorEdPsych.

Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, *62*, 384-414.

Biber, D. (1988). *Variation across speech and writing*. Cambridge, MA: Cambridge University Press.

Biber, D. (1992). On the complexity of discourse complexity: A multidimensional analysis. *Discourse Process*, *15*, 133-163.

Biber, D., & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text*, *93*, 124.

Bieri, J., Atkins, A. L., Briar, S., Leaman, R. L., Miller, H., & Tripodi, T. (1966). *Clinical and social judgment.* New York, NY: Wiley.

Brajer, V., & Gill, A. (2010). Yakity-yak: Who talks back? An email experiment. *Social Science Quarterly*, *91*, 1007-1024.

Brint, S., Proctor, K., Murphy, S. P., Turk-Bicakci, L., & Hanneman, R. A. (2009). General education models: Continuity and change in the u.s. undergraduate curriculum, 1975-2000. *The Journal of Higher Education*, *80*, 605-642.

Buchner, A., Erdfelder, E., Faul, F., & Lang, A. (2009). *G*power (version 3.0.10)[computer program].* Retrieved from \url{http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/}

Burleson, B. R., & Caplan, S. E. (1988). Cognitive complexity. In J. C. McCroskey, J. A. Daly, M. M. Martin, & M. J. Beatty (Eds.), *Communication and personality: Trait perspectives* (p. 233). Creskill, NJ: Hampton Press.

Burt, C. (1949). The structure of the mind: A review of the results of factor analysis. *British Journal of Educational Psychology*, *19*, 110-111.

Caffarella, R. S., & Barnett, B. G. (2000). Teaching doctoral students to become scholarly writers: The importance of giving and receiving critiques. *Studies in Higher Education*, *25*, 39-52.

Cameron, J., Nairn, K., & Higgins, J. (2009). Demystifying academic writing: Reflections on emotions, know-how and academic identity. *Journal of Geography in Higher Education*, *33*, 269-284.

Campbell-Sills, L., Cohan, S. L., & Stein, M. B. (2006). Relationship of resilience to personality, coping, and psychiatric symptoms in young adults. *Behavior Research and Therapy*, *44*, 585-599.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytical studies.* New York, NY: Cambridge University Press.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, *38*, 592.

Cattell, R. B. (1946). *The description and measurement of personality.* New York, NY: World Book Company.

Cherry, R. D., & Meyer, P. R. (1993). Reliability issues in holistic assessment. In M. M. Williamson & B. A. Huot (Eds.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations* (pp. 109–141). Creskill, NJ: Hampton.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

College Board. (2003). *Accuplacer online technical manual review.* Retrieved from http://professionals.collegeboard.com/profdownload/accuplacer-program-manual.pdf.

College Board. (2022). *SAT essay prompts.* Message posted to http://professionals.collegeboard.com/testing/sat-reasoning/prep/essay-prompts.

Costa, P. T., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO personality inventory. *Personality and Individual Differences*, *12*, 887-898.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.

Davis, D., & Brock, T. C. (1975). Use of first-person pronouns as a function of increased objective self-awareness and performance feedback. *Journal of Experimental Social Psychology*, *11*, 381-388.

Deary, I. J. (2001). Individual differences in cognition: British contributions over a century. *British Journal of Psychology*, *92*, 217-237.

Deary, I. J., & Batty, G. D. (2006). Commentary: Pre-morbid iq and later health—the rapidly evolving field of cognitive epidemiology. *International Journal of Epidemiology*, *35*, 670-672.

Delia, J. G., O'Keefe, B. J., & O'Keefe, D. J. (1982). The constructivist approach to communication. In F. E. X. Dance (Ed.), *Comparative human communication theory* (pp. 147–191). New York: Harper & Row.

del Re, A. C. (2013). *Compute.es: Compute effect sizes.* http://cran.r-project.org/web/packages/compute.es.

Derrida, J. (1976). *Of grammatology* (1st American ed.). Baltimore, MD: Johns Hopkins University Press. (Original work published 1967)

Dixon, W. E. (2003). *Twenty studies that revolutionized child psychology.* Upper Saddle River, NJ: Prentice Hall.

Dunn, T. J., Baguley, T., & Brundsen, V. (2013). From alpha to omega: A practical solution to the pervavise problem of internal consistency estimation. *British Journal of Psychology*, *3*.

Emig, J. (1971). *The composing processes of twelfth graders.* Urbana, IL: National Council of Teachers of English.

Enders, C. K. (2011). *Applied missing data analysis.* New York, NY: Guilford.

Eysenck, H. J. (1976). *The measurement of personality.* Lancaster, England: MTP.

Eysenck, H. J. (1979). *The structure and measurement of intelligence.* New York, NY: Springer-Verlag.

Fast, L. A., & Funder, D. C. (2010). Gender differences in the correlates of self-referent word use: Authority, entitlement, and depressive symptoms. *Journal of Personality*, *78*, 313-338.

Felmlee, D. H. (1999). Social norms in same- and cross-gender friendships. *Social Psychology Quarterly*, *62*, 53-67.

Flowers, L., & Hayes, J. R. (1980). Identifying the organization of writing processes. In L. Gregg & E. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Hillsdale, NJ: Erlbaum.

Foltz, P. W., Kintsch, W., & Landauer, T. K. (2009). The measurement of textual coherence with latent semantic analysis. *Discourse Process*, *25*, 285-307.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*, 2-18.

Furnham, A., Monsen, J., & Ahmetoglu, G. (2009). Typical intellectual engagement, big five personality traits, approaches to learning and cognitive ability predictors of academic performance. *British Journal of Educational Psychology*, *79*, 769-782.

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, *17*, 1-13.

Gilbert, C. K., & Heller, D. E. (2013). Access, equity, and community colleges: The truman commission and federal higher education policy from 1947 to 2011. *Journal of Higher Education*, *84*, 417-443.

Ginsburg, H., Ogletree, J., Silakowski, S. M., & Tammy, D. (2003). Vulgar language: Review of sex differences in usage, attributions, and pathologies. *North American Journal of Psychology*, *5*(1), 105-115.

Gleser, G. C., Gottschalk, L. A., & Watkins, J. (1959). The relationship of sex and intelligence to choice of words: A normative study of verbal behavior. *Journal of Clinical Psychology*, *15*, 182-191.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84-96.

Gottfredson, L. S. (1997). Intelligence and social policy. *Intelligence*, *24*, special issue.

Gottfredson, L. S. (2002). g: Highly general and highly practical. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general intelligence factor: How general is it?* (pp. 331–380). Mahwah, NJ: Erlbaum.

Gottschalk, L. A., Gleser, G. C., & Hambridge, G. C. (1957). The unobtrusive measurement of psychological states and traits. In C. W. Roberts (Ed.), *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts* (pp. 117–129). Mahwah, NJ: Erlbaum.

Graesser, A. C., & McNamara, D. S. (2012). Automated analysis of essays and open-ended verbal responses. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, vol 1: Foundations, planning, measures, and psychometrics* (pp. 307–325). Washington, DC: American Psychological Association.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? some practical clarifications of multiple imputation theory. *Prevention Science*, *8*, 206-213.

Guilford, J. S., & Zimmerman, W. S. (1976). *The Guilford–Zimmerman.*

Gustafsson, J. E. (2001). On the hierarchical structure of ability and personality. In J. M. Collis & S. Messick (Eds.), *Intelligence and personality: Bridging the gap in theory and measurement* (pp. 25–42). Mahwah, NJ: Erlbaum.

Hall, N., Hladkyj, C., Perry, R. P., & Ruthig, J. C. (2004). The role of attributional retraining and elaborative learning in college students' academic development. *Journal of Social Psychology*, *144*, 591-612.

Hall, N. C., Perry, R. P., Goetz, T., J. C. Ruthig, R. S., & Newall, N. E. (2007). Attributional retraining and elaborative learning: Improving academic development through writing-based interventions. *Learning & Individual Differences*, *17*, 280-290.

Hart, R. P. (2001). Redeveloping DICTION: Theoretical considerations. In M. D. West (Ed.), *Theory, method, and practice in computer content analysis* (pp. 43–60). New York, NY: Springer.

Heidegger, M. (1950). Language. In V. B. Leitch (Ed.), *The norton anthology of theory and criticism* (1st ed., pp. 1121–1134). New York: W.W. Norton & Company.

Hofstadter, A. (1971). *Poetry, language and being.* New York: Harper & Row.

Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities.* Unpublished doctoral dissertation.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*, 253-270.

Hunt, K. W. (1964). *Differences in grammatical structures written at three grade levels* (Tech. Rep. No. 1988 Cooperative Research Project). Florida State University.

Ickes, W., Redihead, S., & Patterson, M. L. (1986). Machiavellianism and self-monitoring: As different as 'me' and 'you'. *Social Cognition*, *4*, 58-74.

Ingram, R. E. (1990). Self-focused attention in clinical disorders: Review and a conceptual model. *Psychological Bulletin*, *107*, 156-176.

International Reading Association and National Council of Teachers of English. (1996). *Standards for the english language arts.* Urbana, IL: National Council of Teachers of English. http://www.ncte.org/library/NCTEFiles/Resources/Books/Sample/StandardsDoc.pdf.

Jamshidian, M., & Jalal, S. (2010). Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika*, *75*, 649-674.

Jamshidian, M., Jalal, S., & Jansen, C. (2013). *Missmech: Testing homoscedasticity, multivariate normality, and missing completely at random.* Retrieved from \url{http://CRAN.R-project.org/package=MissMech}

Jelicic, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental psychology*, *45*, 1195-1199.

Jensen, A. R. (1992). Perspectives in individual differences. In A. Glover & R. R. Ronnong (Eds.), *Historical foundations of educational psychology* (pp. 61–88). New York, NY: Plenum Press.

Johnson, R. L., Penny, J., & Belita, G. (2000). The relation between score resolution methods and interrater reliability: An empirical study of an analytic scoring rubric. *Applied Measurement in Education*, *13*, 121-138.

Kelly, G. A. (1955). *The psychology of personal constructs (2 vols.).* New York, NY: W.W. Norton.

Kobrin, J. L., Deng, H., & Shaw, E. J. (2011). The association between SAT prompt characteristics, response features, and essay scores. *Assessing Writing*, *16*, 154-169.

Lakoff, R. T. (1975). *Language and woman's place: Text and commentaries* (2nd ed.). New York, NY: Oxford University Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211-240.

Landauer, T. K., Foltz, P. W., & Latham, D. (1998). Introduction to latent semantic analysis. *Discourse Process*, *25*, 259-284. Retrieved from http://lsa.colorado.edu/papers/dp1.LSAintro.pdf.

Leaper, C., & Robnett, R. D. (2011). Women are more likely than men to use tentative language, aren't they? A meta-analysis testing for gender differences and moderators. *Psychology of Women Quarterly*, *35*, 129-142.

Lewis, E. (2010). Academic writing and tacit knowledge. *Teaching in Higher Education*, *15*, 151-160.

Lew-Williams, C., & Fernald, A. (2007). Young children learning spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*, 193-198.

Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching* (9th ed.). Upper Saddle River, NJ: Pearson.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*, 1198-1202.

Little, R. J. A. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association*, *87*, 1227-1237. doi: 10.1080/01621459.1992.10476282

Lorenz, K., & Cobb, S. (1953). Language patterns in psychotic and psychoneurotic subjects. *Archives of Neurology and Psychiatry*, *69*, 684-693.

Maclellan, E. (2004). How reflective is the academic essay? *Studies in Higher Education*, *29*, 75-89.

Mateos, M., & Solé, I. (2009). Synthesizing information from various texts: A study of procedures and products at different educational levels. *European Journal of Psychology of Education*, *24*, 435-451.

Mattern, K. D., & Packman, S. (2009). *Predictive validity of Accuplacer scores for course placement: A meta-analysis* (Tech. Rep. No. 2009-2). New York, NY: College Board.

McCrae, R. R. (1993). Moderated analyses of longitudinal personality stability. *Journal of Personality and Social Psychology*, *65*, 577-585.

McCrae, R. R., & Costa, P. T. (1986). Personality, coping, and coping effectiveness in an adult sample. *Journal of Personality*, *54*, 385-405.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*, 175-215.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Erlbaum.

McGrew, K. S. (2005). The Cattell-Horn-Carroll (CHC) theory of cognitive abilities. Past, present and future. In D. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 135–202). New York, NY: Guildford Press.

Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of personality and social psychology*, *84*, 857-870.

National Center for Educational Statistics. (2013). *Undergraduate enrollment.* Retrieved from http://nces.ed.gov/programs/coe/indicator_cha.asp.

Nolen-Hoeksema, S. (1987). Sex differences in unipolar depression: Evidence and theory. *Psychological Bulletin*, *101*, 259-282.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, *66*, 574-583.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw–Hill.

Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: problems with using long words needlessly. *Applied Cognitive Psychology*, *20*, 139-156.

Owens, C., & Fiona, A. (2013). A 5-year systematic strategy to reduce plagiarism among first-year psychology university students. *Australian Journal of Psychology*, *65*(1), 14-21.

Pennebaker, J. W. (1993). Putting stress into words: Health, linguistic, and therapeutic implications. *Behavior Research and Therapy*, *31*, 539–548.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007a). *The Development and Psychometric Properties of LIWC2007.* Austin, TX: LIWC.net.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007b). *Linguistic Inquiry and Word Count (LIWC) 2007.* Austin, TX: University of Texas.

Pennebaker, J. W., Groom, C. J., Loew, D., & Dabbs, J. M. (2004). Testosterone as a social inhibitor: Two case studies of the effect of testosterone treatment on language. *Journal of Abnormal Psychology*, *113*, 172-175.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*, 1296-1312.

Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the life span. *Journal of Personality and Social Psychology*, *85*, 291-301.

Pfister, F. R., & Petrik, J. F. (1980). A heuristic model for creating a writer's audience. *College Composition and Communication*, *31*, 213-220.

Porchea, S., Allen, F., Robbins, J., & Phelps, R. P. (2010). Predictors of long-term enrollment and degree outcomes for community college students: Integrating academic, psychosocial, socio-demographic, and situational factors. *Journal of Higher Education*, *81*, 680-708.

Provasnik, S., & Planty, M. (2008). *Community colleges: Special supplement to the condition of education 2008*. Washington, DC: National Center for Education Statistics.

Rademaker, A. R., van Zuiden, M., Vermetten, E., & Geuze, E. (2011). Type d personality and the development of ptsd symptoms: A prospective study. *Journal of Abnormal Psychology*, *120*, 299-307.

R Core Team. (2013). *R: A language and environment for statistical computing.* Retrieved from \url{http://www.R-project.org/}

Revelle, W. (2014). *Package 'psych'.* Retrieved from \url{https://personality-project.org/r/psych.manual.pdf}

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega and the glb: Comments on Sijtsma. *Psychometrika*, *74*, 145-154.

Sanders, J. M., & Schilperoord, J. (2006). Text structure as a window on the cognition of writing: How text analysis provides insights in writing products and writing processes. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 1–2). New York: The Guildford Press.

Sanford, F. H. (1948). Speech and personality. In L. A. Pennington & I. A. Berg (Eds.), (pp. 157–177). London, England: Oxford.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Thousand Oaks, CA: Sage.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology*, *57*, 1-10. doi: 10.1037/a0018082

Schnurr, P. P., Rosenberg, S. D., Oxman, T. E., & Tucker, G. J. (1986). A methodological note on content analysis: Estimates of reliability. *Journal of Personality Assessment*, *50*, 601-609. Retrieved from the Psychology and Social Sciences Collection database.

Seligman, M. E. P., Peterson, C., Schulman, P., & Castellon, C. (1992). The explanatory style scoring manual. In C. P. Smith (Ed.), *Motivation and personality: Handbook of thematic content analysis* (p. 383). Cambridge, MA: Cambridge University Press.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*, 107-120.

Spearman, C. E. (1904). 'General intelligence,' objectively determined and measured. *American Journal of Psychology*, *15*, 201-292.

Spearman, C. E. (1927). *The abilities of man.* London, England: Macmillan and Company.

Stein, N. L., Folkman, S., & Richards, T. A. (1997). Appraisal and goal processes as predictors of psychological well-being in bereaved caregivers. *Journal of Personal and Social Psychology*, *72*, 872-874.

Stirman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, *63*, 517-522. Retrieved from the HighWire Press database.

Suedfeld, P., Tetlock, P. E., & Streufert, S. (1992). Conceptual/integrative complexity. In C. P. Smith (Ed.), *Motivation and personality: Handbook of thematic content analysis* (pp. 393–400). Cambridge, MA: Cambridge University Press.

Thakerar, J. N., Giles, H., & Cheshire, J. (1982). Psychological and linguistic parameters of speech accommodation theory. In C. Fraser & K. R. Scherer (Eds.), *Advances in the social psychology of language* (pp. 205–255). Cambridge, MA: Cambridge University Press.

Trapnell, P. D., & Campbell, J. D. (1999). Private self-consciousness and the five-factor model of personality: Distinguishing rumination from reflection. *Journal of Personality and Social Psychology*, *76*, 284-304.

Vallotton, C., & Ayoub, C. (2011). Use your words: The role of language in the development of toddlers' self-regulation. *Early Childhood Research Quarterly*, *26*, 169-181.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in **R**. *Journal of Statistical Software*, *45*(3), 1-67. Retrieved from http://www.jstatsoft.org/v45/i03.

van Heugten, M., & Shi, R. (2009). French-learning toddlers use gender information on determiners during word recognition. *Developmental Science*, *12*(3), 419-425.

Vygotsky, L. (1986). *Thought and language*. Cambridge, MA: MIT Press. (Original work published 1934)

Weber, R. (2006). Function words in the prosody of fluent reading. *Journal of Research in Reading*, *29*, 258-269.

Weintraub, W. (1989). *Verbal behavior in everyday life*. New York, NY: Springer Publishing Company.

Woodcock, R. W., Mather, N., & McGrew, K. S. (2001). *Woodcock-johnson iii tests of cognitive abilities examiner's manual.* Itasca, IL: Riverside Publishing.

Yi-Tai, S., Chung, C. K., & Pennebaker, J. W. (2011). Experimental manipulations of perspective taking and perspective switching in expressive writing. *Cognition and Emotion*, *25*, 926-938.