

# GRE<sup>®</sup>

RESEARCH

## GRE<sup>®</sup> Automated-Editing-Task Phase II Report: Item Analyses, Revisions, Validity Study, and Taxonomy Development

Hunter Breland  
Karen Kukich  
Lisa Hemat

May 2001

GRE Board Professional Report No. 98-05P

ETS Research Report 01-12



Princeton, NJ 08541

GRE® Automated-Editing-Task Phase II Report:  
Item Analyses, Revisions, Validity Study, and Taxonomy Development

Hunter Breland  
Karen Kukich  
Lisa Hemat

GRE Board Report No. 98-05P

May 2001

This report presents the findings of a  
research project funded by and carried  
out under the auspices of the  
Graduate Record Examinations Board.

Educational Testing Service, Princeton, NJ 08541

\*\*\*\*\*

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in Graduate Record Examinations Board reports do not necessarily represent official Graduate Record Examinations Board position or policy.

\*\*\*\*\*

The Graduate Record Examinations Board and Educational Testing Service are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.

EDUCATIONAL TESTING SERVICE, ETS, the ETS logos, GRADUATE RECORD EXAMINATIONS, and GRE are registered trademarks of Educational Testing Service. MCAT® is a registered trademark of the Association of American Medical Colleges. GMAT® is a registered trademark of the Graduate Management Admission Council.

Educational Testing Service  
Princeton, NJ 08541

Copyright © 2001 by Educational Testing Service. All rights reserved.

## Abstract

Two automated editing tasks developed in a Phase I study were subjected to item analyses, revised, and then used in a computer-based test administration at a local college. The data collected in the administration were compared with questionnaire data obtained from students to examine the construct validity of the tasks. In a second approach to construct validation, a taxonomy of writing skills was developed and compared to the skills assessed by the editing tasks. Data analyses indicate that total editing score correlates more strongly with self-reported English grades than with self-reported mathematics grades, and that total editing score correlates positively with student self-assessments of their writing skill, recent grades on writing assignments, and college grade point average. A review of the task elements against the taxonomy indicates that the editing tasks assess important writing skills not assessed by free-response essays. For this and other reasons, it was concluded that automated editing tasks would serve as a useful complement to free-response writing assessments.

Key words: Writing, assessment, automation, essays, diagnosis

## **Acknowledgements**

The editing tasks analyzed in this study were originally developed by Hilary Herbold of the Princeton University English Department. Over the time that the tasks have been under development, however, numerous changes have been made by others. In Phase II of the project, Diane Gruenberg of The College of New Jersey reviewed and modified the tasks based on item analysis results. Probal Tahbaldar programmed the tasks for Phase I of the project and revised the tasks for Phase II. Recognition is also due to the Department of Psychology of the College of New Jersey for use of their computer laboratory, and to Isaac Bejar, Claudia Leacock, and Lawrence Stricker for helpful reviews of an earlier version of the report.

## Table of Contents

	Page
Introduction .....	1
Problems in Writing-Skill Assessment .....	2
Reliability .....	3
Efficiency .....	3
Construct Representation .....	3
Diagnosis .....	4
Authenticity .....	5
Prompt Comparability .....	5
Automated Approaches to Writing Assessment .....	6
Method .....	8
Instruments .....	8
Sample and Data Collection .....	10
Data Analyses .....	10
Taxonomy Development .....	11
Results .....	13
Data Description .....	13
Correlational Analyses .....	14
Item Analysis Results .....	16
Taxonomy Analysis Results .....	16
Discussion .....	18
Conclusion .....	20
References .....	23
Tables and Figures .....	27
Appendix A .....	35
Appendix B .....	45
Appendix C .....	53

### List of Tables

	Page
Table 1. Data Description, Categorical Variables .....	28
Table 2. Data Description, Continuous Variables .....	29
Table 3. Correlation Matrix, Excluding SAT Scores .....	30
Table 4. Correlation Matrix, Including SAT Scores .....	31
Table 5. Correlations Between First and Second Editing Task Scores and Other Variables .....	32
Table 6. Item Analysis of Phase I and Phase II Editing Tasks .....	33
Table 7. Taxonomy Analysis Results .....	34

### List of Figures

	Page
Figure 1. A model of writing proposed by Hayes and Flower (1980; used by permission): Planning, text generation, and revision .....	27

## Introduction

Recent trends in assessment of writing skill show a distinct movement toward the use of more constructed-response tasks. Testing programs that have added free-response essay assessments in recent years include the Medical College Admission Test (MCAT<sup>®</sup>), the Graduate Management Admission Test (GMAT<sup>®</sup>), and the Graduate Record Examinations (GRE<sup>®</sup>). The domain of constructed-response tasks extends far beyond free-response essay assessments, however, as Bennett (1993) observed:

Even a cursory review of the literature makes clear that constructed response implies a broad range of tasks from relatively minor variations of multiple choice to extended projects and complex performances. For example, as a measure of science achievement the term includes such tasks as inserting the missing word in a sentence about the scientific method, writing an essay describing an experiment, or actually conducting an investigation. A general definition that reflects this diversity is that a constructed response is any question requiring the examinee to generate an answer rather than select from a small set of options. (p. 1-2)

Both Bennett (1993) and Snow (1993) have developed organizational schemes that describe a continuum of constructed-response tasks extending from multiple-choice items to complex presentations and demonstration projects. In between these extremes are a large number of possibilities, including sentence-completion items, corrections of faulty computer programs, and corrections of mathematical expressions. It is within this large, intermediate domain of constructed-response tasks that automated editing resides. When editing tasks involve an entire passage, as in the current study, their complexity extends far beyond that of sentence correction.

This project evolved from a type of writing-skill assessment used by the College Board in the 1960s, known as an "interlinear exercise." It consisted of a double-spaced passage with errors in grammar and sentence structure. The examinee's task in completing the exercise was to identify and correct errors by deleting words and phrases and writing new words and phrases in between double-spaced lines of text. Completed exercises were then scored by trained readers using lengthy scoring guidelines that included examples of appropriate and inappropriate corrections. Scoring was laborious, and required much reader time, but the reliability and validity



of the test were excellent. Nevertheless, use of the test was discontinued because of its cost and inefficiency.

With the advent of computer-based testing, the idea of reinventing the interlinear exercise surfaced in the early 1990s, and some preliminary efforts were made to explore this possibility. These early efforts began by taking two of the old exercises and programming them for computer administration. After a few stops and starts, some programming problems, and some reconceptualizations, it finally became possible to administer exercises of this type by computer. Breland (1998, 1999) describes previous efforts in this area; the next section of this report describes the latest effort.

### **Problems in Writing-Skill Assessment**

Current problems in writing-skill assessment fall into the following categories: reliability, efficiency, construct representation, diagnosis, authenticity, and prompt comparability. Reliability problems stem from the use of only one or two writing tasks in assessments because of testing-time limitations. However, reliability can be enhanced through the use of a complementary multiple-choice test of grammar and sentence structure problems. Efficiency problems occur at both the test development stage and at the scoring stage: the development of good writing prompts is time consuming, and scoring usually requires at least two trained human readers.

Construct representation problems occur primarily because of testing-time limitations – because only first-draft writing is possible within the time available, for example. Since writing responses are most often scored holistically, diagnosis is usually not provided; analytical scoring is usually considered to be excessively labor intensive. Writing assessments are sometimes viewed as nonauthentic because they do not replicate real-life writing tasks. Prompt comparability problems result from the use of only one or two writing prompts, which are not always of equal difficulty for different prompt assignment groups, gender groups, language groups, or ethnic groups. Each of these writing assessment problems is discussed in more detail in the following paragraphs.

## Reliability

The reliability of writing-assessment scores has been studied for many years. One of the first careful studies was reported by Coffman (1966) and indicated that the score reliability of assessments consisting of single essay tasks was less than .50 even when two readers were used. A later replication of this early study by Breland, Camp, Jones, Rock, and Morris (1987) yielded slightly higher estimates of score reliability for single essay tasks: .47 to .64 with two readers, depending on writing task type and computational method. The reliability of current, two-essay writing assessments, such as those used by the GMAT and GRE exams, is nominally about .70 (see Schaeffer, Briel, & Fowles, 2001), although a reliability of .80 is often considered a minimum figure for any high-stakes educational assessment (e.g., see Reckase, 1995). Score reliability should not be confused with interrater reliability, which is only concerned with agreement among readers and does not recognize assessment error resulting from the use of only one or two writing tasks.

## Efficiency

The development of essay assessments of writing skill is relatively labor intensive. The prompts used must be carefully pretested and then evaluated by human readers to ensure that examinees will be able to respond adequately. Additionally, a large number of prompts are needed to prevent examinees from remembering them and informing subsequent examinees of the topics – a problem exacerbated by computer-based testing because test administrations are continuous. Efficiency is also a problem in the scoring of writing assessments, since readers are often convened in a group for instruction and scoring. The scoring, usually by two different readers, is costly and time-consuming, and results in a delay in reporting scores to examinees.

## Construct Representation

When the time available for writing-skill assessment is limited, as it often is, free-response writing is usually limited to first-draft writing, even though it is well known that there is more to writing than that. Models of writing skill developed over a number of years through protocol analysis suggest that revision skills are just as important as drafting skills (Hayes &

Flower, 1980; Hayes, 1996). The model of writing proposed by Hayes and Flower (see Figure 1) illustrates three cognitive writing processes – planning, text generation, and revision – and suggests that each is of equal importance. First-draft writing, such as that generated in GMAT and GRE writing assessments, represents at best only two of three major components of writing – planning and text generation – and the time for planning is very limited. Clearly, such writing tasks do not represent the full domain of writing skills as defined by Hayes and Flower.

### Diagnosis

The GMAT Analytical Writing Assessment (AWA), which is similar to the new GRE Writing Assessment, has been criticized by Rogers and Rymer (1995a, 1995b) because it provides no diagnostic information useful in placing or advising students. Holistic scoring has a long and successful record as one effective method of writing-skill assessment, but even its strongest advocates recognize its limitations as a diagnostic tool (White, 1993). Even though the scoring guide used for the GMAT AWA designates the characteristics of essays at each score level, essays receiving any single holistic score may manifest different writing deficiencies. For example, a score of 3 on the GMAT holistic scale is assigned for papers written in response to the "Analysis of an Issue task" that exhibit one or more of the following characteristics (Graduate Management Admission Council, 1994, pp. 350-357):

- is vague or limited in developing a position on the issue
- is poorly organized
- is weak in the use of relevant reasons or examples
- uses language imprecisely and/or lacks sentence variety
- contains occasional major errors or frequent minor errors in grammar, usage, and mechanics

Thus, even when a holistic scoring guide designates the characteristics associated with essays at each score level, a given holistic score may be assigned for various reasons. Thus, a score does not necessarily indicate what kind of training an examinee may need. It would be useful to know, for example, if the examinee needs training in the conventions of standard

written English. Conversely, it would be useful to know that, even though an examinee appeared to know the conventions of standard written English, he or she was still unable to respond well to the writing prompt. Such information would be useful in making admissions decisions, as well as in making decisions about instruction or placement in the event an examinee is admitted.

### Authenticity

A number of influential educational writers have called for more authenticity in educational assessment (e.g., see Resnick & Resnick, 1990; Wiggins, 1989, 1993), although Dwyer (1993) has observed that there is considerable confusion about just what authentic assessment is perceived to be. Usually, authentic assessment is considered to have the following qualities:

- It is direct, rather than indirect.
- It is composed of dependent, related parcels of knowledge, rather than decomposed into independent pieces of knowledge.
- It is contextualized, rather than assuming that competence can be assessed in a different context from that in which it is practiced and used.
- It most often requires human judges to evaluate materials produced by examinees.

In some senses, therefore, writing assessments that require examinees to write and are evaluated by human judges are more authentic than multiple-choice assessments of writing knowledge. Nevertheless, the brief time usually allowed for writing assessments does not satisfy all proponents of authenticity. Wiggins (1993), for example, asserts:

Thus whatever assessors are testing in a 20-minute essay, it is certainly not the ability to write. As those of us who write for a living know, writing is revision, a constant returning to the basic questions of audience and purpose ... . (208)

### Prompt Comparability

Whenever only one or two different writing prompts are used to assess writing skill, it is important that all examinees be assigned prompts of equivalent difficulty. It is also important that the prompts be of equivalent difficulty for all groups of examinees, such as those defined by

gender or race/ethnicity. Since no entirely satisfactory methods of equating writing prompts exist, careful examination of the performance of different groups of examinees becomes especially important. An example of this kind of work is the study of computer-based prompts for the Test of English as a Foreign Language (TOEFL®) by Breland, Muraki, and Lee (2000), in which the performance of prompt assignment groups (groups of examinees assigned different prompts by the computer algorithm) and gender groups were compared, and prompts were identified for possible replacement or revision.

### **Automated Approaches to Writing Assessment**

One important approach to resolving current problems in writing-skill assessment involves the use of natural language processing (NLP) to assist in the scoring and diagnosis of examinee responses to essay prompts. NLP has already been successful in improving the efficiency of writing-skill assessments by reducing the number of readings of examinee responses that are needed (see Kukich, 2000). NLP is also likely to be useful in the diagnosis of examinee responses.

Automated editing represents another important approach toward resolving current problems. Automated editing is intended to be primarily responsive to the problems of reliability, efficiency, and construct representation, although it may also contribute to resolving problems in diagnosis and authenticity. For example, the replacement of one of the two essays now used for both the GMAT and GRE writing assessments with an automated editing task could possibly increase the reliability of these assessments beyond .80, which would approach the reliability of the SAT® II Writing assessment.<sup>1</sup> The GMAT and GRE writing assessments currently have score reliabilities of about .70, so that the replacement of one of the essays with an editing task that has an estimated reliability of about .84 would likely increase the overall score reliability of the assessment.

---

<sup>1</sup> The SAT II writing assessment consists of one 20-minute essay and a 40-minute multiple-choice test of writing, with an estimated reliability for the composite score of .84 (see Willingham & Cole, 1997, p. 282f).

Automated editing is also very efficient. The tasks developed so far have been completed by examinees in an average of 26 minutes, and scoring is immediate. Moreover, scores can be easily copied to a database for analysis and reporting.

Perhaps the most important contribution that editing tasks can make is in the area of construct representation. Since first-draft writing tasks allow little time for editing, and since they do not focus on specific editing problems, editing tasks can provide a useful complement to free-response writing tasks. Editing tasks can also point to specific areas of writing skill in which examinees need improvement. In this regard, they have the potential to contribute to the integration of assessment and instruction, which has long been an elusive goal of assessment.

Finally, editing tasks are authentic measures, in that they present the examinee with a real-life problem – a whole piece of writing in need of improvement. Unlike multiple-choice tests of writing, which consist of discrete sentences having errors in grammar and sentence structure, editing tasks contain sentences in context. In this format, problems of many types may be posed – such as problems with transitions, redundancies, contradictions, and non sequiturs, for example. These are the kinds of writing problems many people face in everyday life. Graduate students have a special need for these kinds of writing skills; writing is important to their graduate training, and once they have completed their graduate programs, many will enter professions where the written word will be their daily companion. It would therefore seem to be important for prospective graduate students to demonstrate that they are well on the way to developing these skills.

The current study involves two phases of automated-editing research. During Phase I, two tasks were developed and then programmed for the computer. Following informal tryouts by internal ETS staff, a limited data collection with 52 students was conducted. These students completed the two tasks and a questionnaire asking about their writing experiences, their overall grades, their grades in writing courses, and their perceptions of their writing abilities (as compared to those of their peers). Data analyses were conducted to investigate the amount of time students required to complete the tasks, as well as the reliability and validity of the tasks.

The time students required to complete both tasks averaged 27 minutes, although some students completed them in much less time (as little as 14 minutes) and some took longer (as much as 45 minutes). The reliability of both tasks taken together was estimated at .84. The correlation of the total score for both tasks with students' self-assessments of their writing ability was .52; with overall grades, .14; and with grades in writing courses, .46. (For more information about Phase I work, see Breland, 1998a, 1998b, & 1999.)

Phase II, which is reported here, encompassed several project objectives:

- to conduct item analyses using Phase I data
- to revise the tasks based on the results of the item analyses
- to conduct a paper-and-pencil data collection using the revised tasks
- to revise the tasks a second time based on the paper-and-pencil data collections
- to revise the computer program
- to collect data in a totally electronic environment
- to analyze the data
- to explore the scope of writing skills assessed by the editing tasks

This last objective posed the need for a fairly comprehensive taxonomy of writing skills against which we could compare those skills assessed in the revised tasks. In the final stage of this project, we compiled a preliminary taxonomy of principles of good writing and common grammatical and stylistic flaws to use for this purpose.

## **Method**

### Instruments

The two automated editing tasks used for this study, as well as instructions for completing them and scoring procedures, are provided in Appendix A. The questionnaire subjects completed about their background and writing experiences is provided in Appendix B. This questionnaire was similar to one used earlier by Breland (1998a, 1998b) and developed by

Powers, Fowles, and Boyles (1996). It asked students about their grades overall, their grades in writing courses, their writing accomplishments, their self-perceptions of their writing ability (as compared to that of their peers), and their college admission test scores. The following variables were developed from the questionnaire:

1. Gender (Questionnaire Item 1): Female = 1, Male = 2.
2. Ethnic status (Questionnaire Item 2): Black = 1, American Indian = 2, Asian = 3, White = 4, Hispanic = 6, Other = 7.
3. English dominant (Questionnaire Item 3, Is English the dominant language in your household?): Yes = 1, No = 2.
4. English better (Questionnaire Item 4, Do you understand English as well as or better than any other language?): Yes = 1, No = 2.
5. Writing ability (WABIL, Questionnaire Item 5): Well above average = 5, Somewhat above average = 4, About average = 3, Somewhat below average = 2, Well below average = 1.
6. Writing success in different areas of writing (AREA, Questionnaire Item 6): This variable was developed by averaging responses made on a scale of 1 (not at all successful) to 5 (extremely successful) in four different areas of writing (thinking about an assignment, organizing, revising, and developing an effective writing style).
7. Writing success in different genres (GENRE, Questionnaire Item 7): This variable was developed by averaging responses made on a scale of 1 (not very successful) to 5 (extremely successful) for eight different genres of writing (personal, creative, persuasive, analytical/critical, descriptive/expository, examination, and applied).
8. College grades in math, science, and similar courses (MATHGR, Questionnaire Item 8): This variable was developed by placing grades on a scale of 1 (D or less) to 7 (A).
9. College grades in English, history, and similar courses (ENGGR, Questionnaire Item 8): This variable was developed by placing grades on a scale of 1 (D or less) to 7 (A).
10. Most recent writing grade (RECGR, Questionnaire Item 8): This variable was developed by placing grades on a scale of 1 (D or less) to 7 (A).
11. Writing activities (WACT, Questionnaire Item 9): This variable was developed as a count of activities checked as "Yes," with certain activities double-weighted (for example, "Participated in an essay contest" received one point, while "Won an essay contest" received two points).



12. Status in college (Questionnaire Item 10): Freshman = 1, Sophomore = 2, Junior = 3, Senior = 4.
13. College grade point average (Questionnaire Item 11): This variable was developed using the standard 4-point (0 to 4) scale.
14. Graduate school plans (Questionnaire Item 12, Do you think you will go to graduate school?): Yes = 1, No = 2.
15. SAT I verbal score (SATV, Questionnaire Item 13).
16. SAT I mathematics score (SATM, Questionnaire Item 13).
17. SAT II: Writing test score (SATW, Questionnaire Item 13).

### Sample and Data Collection

Subjects were students enrolled at a local state college, and data collection was conducted in a psychology laboratory of the same college. The computer program was installed in five different computers of the laboratory, and students were paid \$20 each to complete the exercises at the laboratory. Students were recruited for the study with the assistance of professors at the college. Our goal was to recruit at least 100 upper-level students who were planning to attend graduate school. Students responses were scored automatically by the computer program, which also recorded the time it took each student to complete each of the editing tasks.

### Data Analyses

Data analyses were conducted to examine test reliability and discriminant and convergent validity. Reliability was estimated by correlating scores from the two equal-length editing tasks and then applying the Spearman-Brown formula. Discriminant and convergent validity was examined using responses to the background questionnaire. Correlations were computed between scores on the editing tasks and coded responses to questionnaire items.

Item analyses using data collected in Phase I and Phase II were conducted by computing the percentage of subjects who responded correctly to items and by computing biserial correlations (more advanced methods of item analysis were not employed because of data limitations and because no corrections were made for item overlap with total score). Early item

analysis results were used to revise the two editing tasks. Following the task revisions, a paper-and-pencil data collection was conducted to examine the degree to which improvements were made in the items. This time, students at a local college were invited to complete the tasks in a room of the college, and were paid \$20 for their participation. A total of 51 students participated. Following the paper-and-pencil data collection, the tasks were again revised, this time to reflect student solutions deemed acceptable by writing experts. Additionally, revisions were made to incorporate more transitional problems into the tasks.

### Taxonomy Development

An ideal test of editing and revision skills would include a representative sample of the most common flaws in early-draft writing, including both sentence-oriented and discourse-oriented flaws. Sentence-oriented flaws are those that require processing of only one sentence to detect and correct, whereas discourse-oriented flaws require processing of surrounding sentences as well. Errors in usage or in subject-verb agreement are examples of self-contained, sentence-oriented writing flaws. Inappropriate use of logical transition terms (e.g., *furthermore*, *in contrast*) and faulty parallelism across sentences are examples of discourse-oriented writing flaws. Since authentic editing tasks include both types of flaws, a taxonomy of these errors would be useful for assessing editing and revision skills.

One labor-intensive way to derive such a taxonomy would be to compile a corpus of early-draft writings and hand-annotate all of the text segments in need of revision. Theoretically, existing grammar and spelling correction software could aid in the annotation task, but current technology is inadequate for this purpose. Not only are existing grammar-checking programs limited to sentence-oriented errors, but they are also lacking in coverage and accuracy. For example, we submitted the text of the two editing tasks used in this study to one widely used grammar and spelling checker. It correctly detected only seven of more than 80 flaws in the two texts, and it wrongly detected three invalid flaws. The seven valid flaws included two out of five instances of improper punctuation, two out of two instances of *which/that* confusion, one out of three instances of possessive/plural confusion, one sentence fragment, and one occurrence of the word ‘irregardless,’ which is considered to be nonstandard usage. The three flaws wrongly

detected by the grammar checker involved valid uses of the passive voice, of the word *principle*, and of the phrase *but also*.

Ongoing research in NLP techniques for detecting flaws in writing (e.g., see Kukich, 1992; Chodorow & Leacock, 2000; Miltsakaki & Kukich, 2000) could contribute to the task of building a comprehensive taxonomy of writing flaws. Indeed, in the long term, such tools will likely become components of practical authoring assistance software. However, for the purposes of this study, we opted to compile a preliminary taxonomy of writing flaws using traditional sources and methods.

We surveyed the literature and then selected four diverse sources (Strunk & White, 2000; Williams, 1997; Gopen & Swan, 1990; and ETS, 1998), from which we gathered and organized writing principles and errors. We focused on these four sources for the following reasons. *The Elements of Style* (Strunk & White, 2000) is considered a time-tested classic. *Ten Lessons in Clarity and Grace* (Williams, 1997) is widely regarded as a new classic in writing instruction circles. *The Science of Scientific Writing* (Gopen & Swan, 1990) is a succinct summary of mainly discourse-oriented writing principles for professional writers. And *Diagnostic Statements from the Diagnostic Writing Service* (ETS, 1999) represents principles and problems deemed important by writing instruction experts. The larger set of sources surveyed included: Bejar (1988), Crowell and Kolba (1999), Flower (1979), Horvath (1984), Klagholz, Schechter, and DeMauro (1998), Nichols (1994), Sommers (1982), and additional unpublished documents.

Our approach was to first extract as many principles of good writing – as well as grammatical and stylistic flaws – from these sources as possible, and then organize them into related categories. This process yielded more than 500 individual writing principles or flaws. Because there was much redundancy and cross-referencing among the 500-plus entries, and because we felt the need to create sentence-oriented and discourse-oriented categories, we went through several iterations of clustering before settling on six main categories of principles and flaws: lexical, grammar, syntax, punctuation, coherence, and style. We then subdivided each of the first five categories into sentence-oriented and discourse-oriented problems, yielding a total of eleven top-level categories, which we further subdivided, yielding a hierarchy of

approximately 48 bottom-level categories. This number is approximate because category boundaries are fuzzy, and it is not always clear whether to collapse two categories into a single one. An outline of the preliminary taxonomy is shown in Appendix C.

## Results

### Data Description

Table 1 describes the sample of students who participated in the data collection in terms of categorical variables. Out of a total of 114 students who participated, 89 (78.1 %) were female and 25 (21.9%) were male. The racial/ethnic representation was 84.2% White, 5.3% Black, 5.3% Asian, and 3.5% Hispanic. English was the dominant spoken language of 93.9% of the students, and 95.6% reported that English was their best language. Most of the students who participated were juniors or seniors (57.9%); only one student was a freshman. A large proportion (80.5%) of the students reported that they planned to go to graduate school.

Table 2 describes the data collected for continuous variables. Of the 114 students who participated, all completed the background questionnaire, all but one reported a college grade point average (GPA), 99 reported SAT I verbal and mathematics scores, and 24 reported SAT II: Writing scores. A small number also reported SAT I combined scores, but these combined scores were not used.

Of the 114 students who participated, 112 provided responses to the first editing task and 111 completed the second editing task. Average scores for the first and second editing tasks were similar (17.56 vs. 17.48), as were the standard deviations (5.52 vs. 5.86). The average time spent on the first task was 14.14 minutes, and that for the second task was 11.98 minutes. However, the time required for reading the instructions was included in the first task time, but not the second. The average total time required to complete both editing tasks was 26.06 minutes, although one student completed both tasks in only 7.03 minutes and another required 51.65 minutes to complete both tasks. Table 2 also shows that some students spent very little time responding to the tasks (as little as 1.15 minutes for the first task and 3.57 minutes for the second task), and it

was obvious from an examination of response patterns that not all students who responded to a task were highly motivated. For this reason, 12 cases were deleted before further data analysis.

### Correlational Analyses

Because the number of cases available for analysis were limited for some variables, especially SAT scores, two sets of correlational analyses were conducted. The first set, which excludes SAT scores, is given in Table 3. Table 3 provides correlations between the total editing score (EDIT) and variables generated from the background questionnaire, as well as intercorrelations among all variables. The highest correlation with total editing score (.46) was for self-reported English, history, and similar grades (ENGGR); this correlation is somewhat higher than that obtained for math, science, and similar grades (MATHGR, .31). Total editing score also correlated with students' self-rating of their own writing ability (WABIL, .35), college GPA (GPA, .35), and the grade received on the most recent writing assignment (RECGR, .29). A somewhat lower correlation was obtained with writing activities and accomplishments (WACT, .16).

These correlations are similar to correlations obtained by Powers, Fowles, and Boyles (1996) between variables developed from a similar questionnaire and essay scores. In that study, "grade average in writing courses" correlated .35 with essay scores, as compared to the .46 correlation obtained in the present study between self-reported "English, history, and similar grades" (ENGGR) and total editing score. However, other correlations obtained in the previous study were higher than those obtained here. For example, students' self-ratings of their writing ability, as compared with their peers, correlated .46 with essay scores in the previous study versus the .35 correlation obtained in the present study between students' self-rating of their own writing ability (WABIL) and total editing score. Correlations larger than .20 in Table 3 are significant at the .05 level, and correlations larger than .25 are significant at the .01 level.

Table 4 provides the correlations obtained when SAT I verbal and mathematics scores were included in the analysis, which reduced the number of cases available to 93. The highest correlation with total editing score was still self-reported English, history, and similar grades

(ENGGR, .44), which was somewhat higher than the correlation with self-reported math, science, and similar grades (MATHGR, .25); this difference in correlations is statistically significant ( $p < .05$ , one-tailed). SAT I verbal and mathematics scores correlated .39 and .36, respectively, with total editing score. Other correlations in Table 4 are generally similar to those shown in Table 3, suggesting that reducing the analysis from 102 cases to 93 cases did not substantially change relationships.

Correlations with SAT I verbal score are similar to correlations obtained by Powers, Fowles, and Boyles (1996) with GRE verbal score. In the previous study, "grade average in writing courses" correlated .36 with GRE verbal score, as compared to the correlation of .44 obtained in the present study between self-reported English, history, and similar grades (ENGGR) and SAT I verbal score. However, the correlation between students' self-ratings of their writing ability (WABIL) and GRE verbal score was .43 in the previous study versus .29 for SAT I verbal score here. In Table 4, correlations larger than .20 are significant at the .05 level, and correlations larger than .27 are significant at the .01 level.

SAT II: Writing scores were reported by only 24 students, and 20 students provided both these scores and usable total editing scores. The correlation between SAT II: Writing score and total editing score was .75, suggesting a good relationship between these two variables, even though the data were limited.

A reliability estimate for total editing score was obtained by correlating scores on the two editing tasks ( $N = 102$ ) and then applying the Spearman-Brown formula. The correlation between the scores on the two tasks was .73, and the reliability of the total score was estimated at .84. The two editing tasks were also reasonably consistent in their correlations with other variables. Table 5 shows that self-reported English, history, and similar grades (ENGGR), the highest correlate of total editing score, correlated .41 with the first editing task (Free Speech) and .45 with the second (Materialistic Desires). The average of these correlations with other variables was .30 with the first task and .28 with the second.

### Item Analysis Results

A number of item analyses and task revisions were conducted, but the item analysis results reported here represent only the final revisions to the tasks used in Phases I and II of this study. The results of the item analyses, organized by task and project phases, are listed in Table 6. For each task and project phase, Table 6 provides the percentage of subjects responding correctly to an item and the biserial correlation. The average of these statistics is given at the bottom of the table. For example, it can be seen that, for Task 1 (Free Speech), the average proportion of subjects responding correctly was .57 in Phase I and .53 in Phase II, and the average biserial correlation for Task 1 was .31 in Phase I and .39 in Phase II. For Task 2 (Materialistic Desires), the average proportion of subjects responding correctly was .53 in Phase I and .44 in Phase II, and the average biserial correlations for Task 2 were .33 in Phase I and .37 in Phase II.

The tasks were revised for Phase II to optimize the average proportion correct at .50 and to increase the average biserial correlation. Phase II results show that these objectives were met, with the exception of the proportion of subjects responding correctly for Task 2, which decreased from .53 in Phase I to .44 in Phase II. Individual item biserial correlations in Table 6 are quite unstable because of the limited number of cases used, so it is inappropriate to attempt precise comparisons of changes in the biserial correlations between Phase I and Phase II for individual items.

### Taxonomy Analysis Results

The primary motivation for compiling the taxonomy of writing skills was to explore the scope of writing errors assessed by the two editing tasks. Each of the two editing tasks used in this study contains 32 errors, or test items, but some items represent more than one opportunity for improvement. In all, writing experts identified and categorized a total of 83 problems or opportunities for improvement within those 64 items. We found that these 83 problems fall into 24 of the low-level categories in the taxonomy; thus the tasks represent 50 percent of the taxonomy.

Table 7 displays a breakdown of the 83 problems by category. In brief, we counted the following errors in top-level categories: 37 lexical, 5 grammar, 18 syntax, 8 punctuation, and 15 coherence. Twenty-five of the 37 lexical editing opportunities fall into "wordiness" or "redundancy" subcategories. Emphasis on this particular type of writing flaw seems warranted given the historical evidence of its pervasiveness in early-draft writing. Another 31 of the editing skills tested by the tasks fall under the grammar, syntax, and punctuation categories. The remaining 15 problems in the two tasks test the student's ability to recognize and correct coherence flaws in writing. Nine of the 83 editing tasks involve discourse-oriented problems; two of these fall into the punctuation category and seven fall into the coherence category. The remaining 74 errors are sentence-oriented.

An example of a discourse-oriented error in the punctuation category is Item 4 in the Free Speech task: "Some claim that order and civility requires limits upon free **[speech; the]** danger of restricting our open expression far outweighs the possible benefit to society generally." An example of a discourse-oriented error in the coherence category is Item 7 in the Materialistic Desires task. This item tests the student's ability to recognize the logical relationship of similarity, as opposed to contrast, between the current sentence and its preceding context. The preceding sentence states, "Human nature desires both progress and recognition, and material goods are tangible symbols of hard work." Item 7 follows with "**[However,]** materialism benefits the culture as well." This item requires that the student recognize the flawed use of the logical transition term *however*, given the context of the previous discourse. The student is expected to change the word *however* to *furthermore*, *moreover*, *also*, *besides*, *in addition*, or *and* to convey the implied logical connection of *similarity*.

Item 25 on the Materialistic Desires task is also a discourse-oriented error: "[**Nevertheless, the**] health of the economy depends upon our desire for material objects." This item requires the student to recognize that the logical relationship between this sentence and the preceding context is one of elaboration rather than contrast. Thus, transition term *nevertheless*, which suggests contrast, is inappropriate and should be replaced by *furthermore*, *in addition*, *finally*, or some other elaboration transition term.



## Discussion

The results obtained in the Phase II data collection are similar to those obtained in the Phase I data collection, although about twice the number of cases were available for Phase II. In Phase I, only 52 cases were available for analysis, and the number of variables was more limited. In Phase II, 102 cases were available for most analyses, but this was reduced to 93 cases when SAT I scores were included.

Both Phase I and Phase II analyses indicate that automated-editing-task scores correlate best with grades in English or similar courses. In Phase II, data were collected both for grades in English, history, and similar courses and for grades in mathematics, science, and similar courses. Phase II analyses show that grades in English, history, and similar courses correlate better with editing-task scores than do grades in mathematics, science, and similar courses. Phase II results also suggest a strong relationship between editing-task scores and SAT II: Writing scores, although data were very limited.

Phase I and Phase II correlational analyses both support the construct validity of the automated editing tasks used in this study. The reliability estimate obtained in Phase II for the total editing score (.84) is identical to that obtained in Phase I. This is about the same level of reliability obtained for multiple-choice tests of writing skill, which use about the same amount of testing time; thus, it would not appear that local dependence resulting from the use of common passages for each of the editing tasks is a serious problem (see Wainer & Thissen, 1996, for a discussion of this problem), because the multiple-choice items are independent. This reliability estimate (.84) is also identical to that which has been reported for the one-hour SAT II: Writing test, which includes a 40-minute multiple-choice component and a 20-minute essay component (see Willingham & Cole, 1997, p. 282f). Finally, the average time required to complete the two editing tasks in Phase II (26 minutes) was very close to that obtained in Phase I (27 minutes), indicating that a subtest consisting of two editing tasks would fit well within the constraints of current educational testing situations.

Psychometrically, therefore, it would appear that an automated editing test made up of two brief editing tasks would be similar in many ways to a multiple-choice test of writing, such

as that used as a part of the SAT II: Writing test. An automated editing task would be different, however, in that examinees would construct their responses without the aid of any hints or suggestions, and guessing would be extremely difficult. Automated editing tasks are also more similar to real-life writing tasks than are multiple-choice tests, and thus they may be viewed as more "authentic."

It seems likely that the use of an automated editing task as a substitute for one of two free-response writing tasks (such as those used in the GRE and GMAT writing assessments) in an assessment would increase the reliability of that assessment. This is likely because combining a test with an estimated reliability of .84 with a lower reliability essay test will result in a greater total reliability than two lower reliability essay tests. Moreover, it would extend the construct of the assessment to include revision, instead of only first-draft writing. A final advantage would be a reduction in the cost of scoring: First, only one free-response writing task would require scoring by human readers. And second, only one set of human scores would be required for use with a NLP scorer, such as *e-rater*, versus the two sets of human scores currently required.

While the primary motivation for compiling a taxonomy was to explore the scope of writing errors assessed by two editing tasks, a well-developed taxonomy could provide many additional benefits. Together with NLP software tools, a taxonomy could facilitate the creation of items for automated edited tasks. For example, the lexical category of the current taxonomy already contains an extensive list of vocabulary terms that frequently result in inappropriate diction and usage problems. NLP software tools could be devised to identify contexts in valid text where errors of these types might be inserted by test developers to create items. NLP tools could also be devised to identify contexts where errors of imprecision could be introduced using the WordNet database (Fellbaum, 1998). For any given noun, verb, adjective, or adverb, WordNet can suggest a variety of related terms – such as synonyms, antonyms, hyponyms,<sup>2</sup> and hypernyms – that could be used imprecisely or inappropriately in place of valid terms in well-written texts. Similarly, NLP tools could be devised to suggest locations in valid texts where grammar, syntax, and punctuation errors might be introduced.

---

<sup>2</sup> A hypernym is a more general synonym (e.g., *canine* is a hypernym for *dog*), whereas a hyponym is a more specific synonym (e.g., *Corgi* is a hyponym for *dog*).

For scoring, many of the same NLP tools could be used to generate potential valid solutions for items. As indicated in the Scoring Procedures section of Appendix A, a thesaurus, preferably automated, can generate synonyms for terms in solution keys that might result in additional valid solutions for items. NLP tools could then check the potential solutions to eliminate those that violate syntactic and lexical collocation constraints. However, a final review by a writing expert would still be required before these additional solutions are added to the keys.

Certain categories of writing flaws remain less tractable for creating and scoring in automatic editing tasks, even with the help of NLP techniques. Most of these writing flaws fall into discourse-oriented categories. The difficulty lies in the fact that such errors span a larger segment of text, as do lengthy subject-verb separation problems, for example, or whole subject and object phrases that place inappropriate stress on given as opposed to new information. In general, the longer the span of text to be changed is, the greater the number of correct or semi-correct variations examinees might generate.

The use of an error taxonomy and NLP tools can also help address authenticity and diagnosis problems in writing assessment. An automated editing task based on a taxonomy of errors that are frequently found in first-draft writing represents an authentic task – one a student is likely to encounter in professional or personal life. In addition, it supplies information needed to identify specific skills and weaknesses inherent in an examinee's pattern of performance.

### **Conclusion**

The results of this study, as well as those of previous studies of automated editing tasks, strongly suggest that the technique has promise. The reliability of .84, estimated from two different sets of subjects in Phase I and Phase II, is about the same as that obtained for multiple-choice tests of writing that require the same amount of testing time. Accordingly, the problem of local dependence often noted for contextualized tasks would appear not to be a problem at all for editing tasks of the type studied.

Both Phase I and Phase II results indicate that automated editing tasks have good discriminant validity in that editing scores correlate better with English grades than with math

and science grades. The Phase II study adds evidence of convergent validity through a correlation with SAT II: Writing subject test scores, and adds construct validity through taxonomic analysis of the writing problems covered.

Automated editing tasks also have some practical advantages. First, the average time to complete two editing tasks in Phase II was only 26 minutes (27 minutes in Phase I), even though subjects were given no time limit and were not rushed in any way. Second, scoring is automatic, so scoring costs are minimized. And third, test data can be easily transferred to other systems for analysis and score reporting.

With the use of automated editing tasks, it becomes possible to envision a large-scale writing assessment that consists of a single free-response essay coupled with a 30-minute editing test, which together yield a score reliability somewhat higher than would be possible using two essays. If the single essay is scored in part by computer, scoring costs are reduced to those of a single reader. As an added advantage, such a test would provide some diagnostic information not provided by essays alone – namely, information on examinees' skill in recognizing and correcting flaws in standard written English. Such information would be especially useful for examinees who do not respond well to the free-response essay topic.

NLP technology can be used to facilitate both the creation and scoring of automated editing assessments. Such tools can be used to locate portions of well-written text that are susceptible to error, and can suggest possible corruptions. NLP tools can also generate lists of alternative corrections for use in automated scoring.

In the longer term, NLP tools can be used to refine the preliminary taxonomy of writing errors that was created manually for this study. As NLP tools become more sophisticated, they will be able to detect a broader range of writing errors in first-draft texts, such as those texts that are produced in current essay exams. This will provide more accurate information about the types and frequencies of errors that naturally occur in first-draft writing. In particular, it will provide insight into the relative proportions of sentence-oriented and discourse-oriented flaws.

One limitation of this particular study resulted from the difficulty of conducting special computer-based test administrations and data collection prior to full implementation. Our modest goal of obtaining at least 100 cases of data was achieved because we were able to use a first-rate computer laboratory of a local institution and we received institutional assistance in recruiting subjects. However, the number of subjects who could participate was necessarily small because of limitations in the numbers of computers available and the time required for test administration.

Finally, it should be emphasized that automated editing is only one example of a large number of possible constructed-response tasks that might be examined in the way we have here. Many other constructed-response tasks can be implemented using computer technology and natural language processing techniques.

## References

- Ackerman, T. A., & Smith, P. L. (1988). A comparison of the information provided by essay, multiple-choice, and free-response writing tests. *Applied Psychological Measurement*, 12(2), 117-128.
- Bejar, I. I. (1988). A sentence-based automated approach to the assessment of writing: A feasibility study. *Machine-Mediated Learning*, 2, 321-332.
- Bennett, R. E. (1993). On the meanings of constructed responses. In R. E. Bennett and W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 1-27). Hillsdale, NJ: Erlbaum.
- Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In J. Miller and P. Eimas (Eds.), *Handbook of perception and cognition: Speech, language, and communication* (Vol. 11, pp. 181-216). San Diego: Academic Press.
- Breland, H. M. (1999). Exploration of an automated editing task as a GRE writing measure (GRE Report No. 96-01; ETS Research Report No. 99-9). Princeton, NJ: Educational Testing Service.
- Breland, H. M. (1998). Writing assessment through automated editing. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego (part of the symposium, "Automated scoring of constructed responses: From construct representation to operational implementation").
- Breland, H. M., Camp, R., Jones, R. J., Rock, D., & Morris, M. (1987). *Assessing writing skill*. New York: College Entrance Examination Board.
- Breland, H. M., Muraki, E., & Lee, Y. (2000, April). Comparability of TOEFL-CBT writing prompts using logistic regression method. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.
- Chodorow, M., and Leacock, C. (2000). An unsupervised method for detecting grammatical errors. In Association for Computational Linguistics (Eds.), *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics* (pp. 140-147). Seattle: Morgan Kaufmann.
- Coffman, W. C. (1966). On the validity of essay tests of achievement. *Journal of Educational Measurement*, 3(2), 151-156.
- Crowell, S. C., & Kolba, E. D. (1999). *The New Jersey GEPA: Grade 8 language arts literacy writing coach*. Montclair, NJ: Educational Design.

Dwyer, C. A. (1993). Innovation and reform: Examples from teacher assessment. In R. E. Bennett & W. C. Ward, Jr. (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 265-289). Hillsdale, NJ: Erlbaum.

Educational Testing Service (1999). Diagnostic statements from the Diagnostic Writing Service. Princeton, NJ: Author.

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Flower, L. (1979, September). Writer-based prose: A cognitive basis for programs in writing. *College English*, 41, 19-37.

Graduate Management Admission Council. (1994). *The official guide for GMAT review* (7th ed.). Princeton, NJ: Educational Testing Service.

Gopen, G. D., & Swan, J. A. (1990). The science of scientific writing: If the reader is to grasp what the writer means, the writer must understand what the reader needs. *American Scientist*, 78, 550-558.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. Michael Levy and Sarah Ransdell (Eds.), *The Science of writing: Theories methods, individual differences, and applications* (pp. 1-27). Mahwah, NJ: Erlbaum.

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg and E. R. Steinberg (Eds.), *Cognitive processes in writing*. Hillsdale, NJ: Erlbaum.

Hobbs, J. (1979). Coherence and coreference. *Cognitive Science*, 3(1), 67-90.

Horvath, B. K. (1984). The components of written response: A practical synthesis of current views. *Rhetoric Review*, 2, 136-145.

Kellogg, R. (1998). Components of working memory in text production. In M. Torrence and G. Jeffery (Eds.), *The cognitive demands of writing: Processing capacity and working memory effects in text production*. Amsterdam: Amsterdam University Press.

Klagholz, L., Schechter, E., & DeMauro, G. E. (1998, February). *Directory of test specifications and sample items for the Elementary School Proficiency Assessment (ESPA), Grade Eight Proficiency Assessment (GEPA), and High School Proficiency Assessment (HSPA) in language arts literacy*. Trenton, NJ: New Jersey State Department of Education.

Kukich, K. (2000, September/October). Beyond automated essay scoring. *IEEE Intelligent Systems*, 15(5), 22-27.

Kukich, K. (1992, December). Techniques for automatically correcting words in text. *ACM Computing Surveys*, 24(4), 377-439.

- Longacre, R. E. (1983). *The grammar of discourse*. New York: Plenum.
- Miller, M. D., & Crocker, L. (1990). Validation methods for direct writing assessment. *Applied Measurement in Education*, 3(3), 285-296.
- Miltsakaki, E., & Kukich, K. (2000). Automated evaluation of coherence in student essays. *Proceedings of the Workshop on Language Resources and Tools in Educational Applications, Linguistic Resources in Education Conference, Athens, Greece*, <http://www.ling.upenn.edu/~elenimi>.
- Nichols, P. D. (1994). A Framework for Developing Cognitively Diagnostic Assessments. *Review of Educational Research*, 64, 575-603.
- Polanyi, L. (1988). A formal model of discourse structure. *Journal of Pragmatics*, 12, 601-638.
- Powers, D. E., Fowles, M. E., & Boyles, K. (1996). *Validating a GRE writing test* (GRE Report No. 93-26B; ETS Research Report No. 96-27). Princeton, NJ: Educational Testing Service.
- Reckase, M. D. (1995). Portfolio assessment: A theoretical estimate of score reliability. *Educational Measurement: Issues and Practice*, 14(1), 12-14, 31.
- Resnick, L. B., & Resnick, D. P. (1990). Tests as standards of achievement in schools. In *The uses of standardized tests in American education: Proceedings of the 1989 ETS Invitational Conference* (pp. 63-80). Princeton, NJ: Educational Testing Service.
- Rogers, P. S., & Rymer, J. (1995a). What is the functional value of the GMAT Analytical Writing Assessment for management education? A critical analysis, Part 2. *Management Communication Quarterly*, 8(4), 477-494.
- Rogers, P. S., & Rymer, J. (1995b). What is the relevance of the GMAT Analytical Writing Assessment for management education? A critical analysis, Part 1. *Management Communication Quarterly*, 8(3), 347-367.
- Schaeffer, G. A., Briel, J. B., & Fowles, M. E. (2001). *Psychometric evaluation of the new GRE Writing Assessment* (GRE Report No. 96-11P). Princeton, NJ: Educational Testing Service.
- Snow, R. E. (1993). Construct validity and constructed-response tests. In R. E. Bennett and W. C. Ward Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 45-60). Hillsdale, NJ: Erlbaum.
- Sommers, N. (1982). Responding to student writing. *College Composition and Communication*, 33, 148-156.



Strunk, W., & White, E. B. (2000). *The elements of style* (4th ed.). Needham Heights, MA: Allyn & Bacon.

Swanson, D. B., Norman, R. R. & Linn, R. L. (1995). Performance-based assessment: Lessons from the health professions. *Educational Researcher*, 24(5), 5-11, 35.

Wainer, H., & Thissen, D. (1996). How is reliability related to the quality of test scores? What is the effect of local dependence on reliability? *Educational Measurement: Issues and Practice*, 15(1), 22-29.

White, E. M. (1993). Holistic scoring: Past triumphs, future challenges. In M. M. Williamson & B. A. Huot (Eds.), *Validating holistic scoring for writing assessments: Theoretical and empirical foundations* (pp. 79-108). Cresskill, NJ: Hampton.

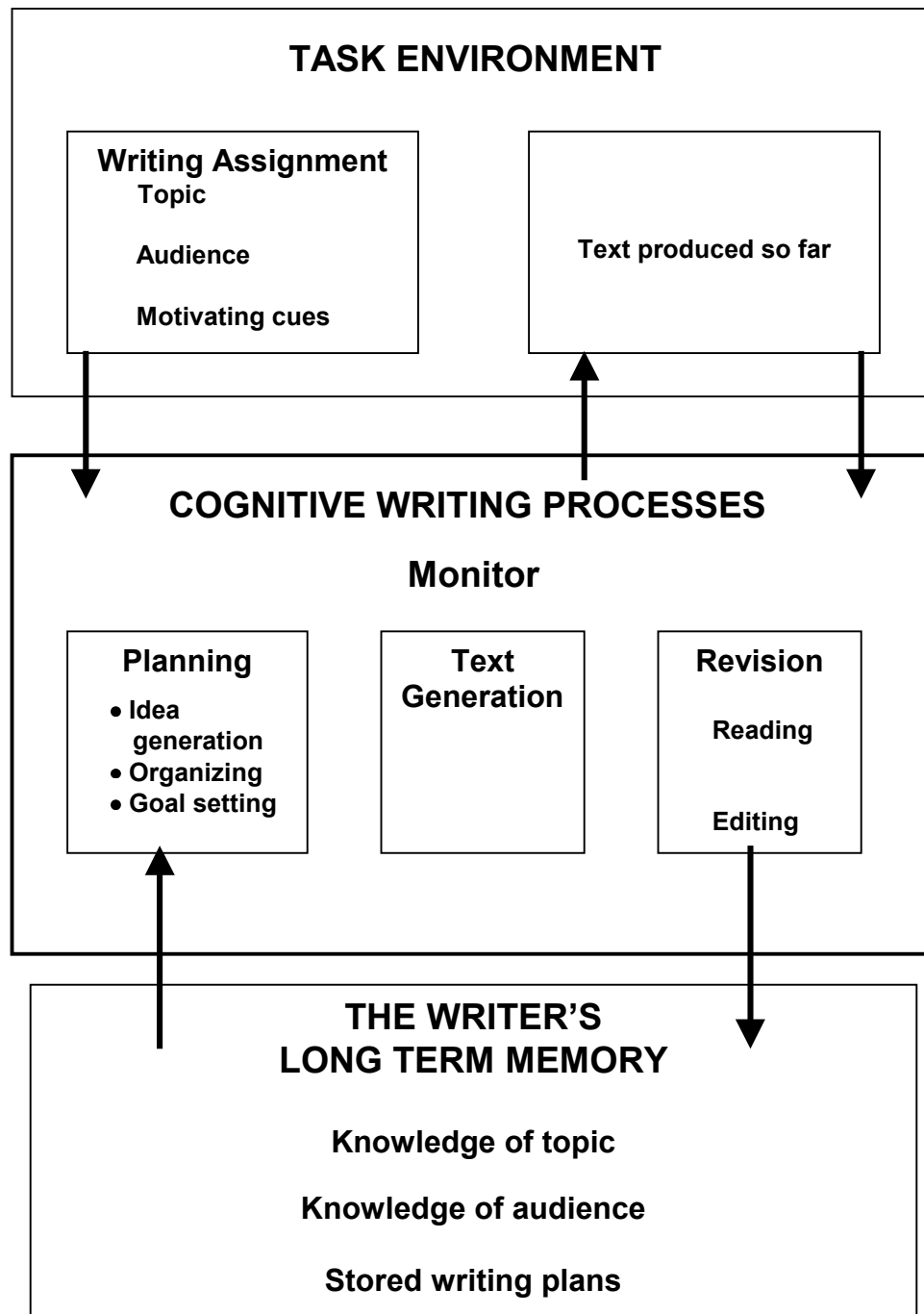
Wiggins, G. (1993). Assessment: Authenticity, context, and validity. *Phi Delta Kappan*, 75, 200-214.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70, 703-713.

Williams, J. M. (1997). *Style: Ten lessons in clarity and grace* (5<sup>th</sup> ed.). New York: Addison Wesley Longman.

Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Erlbaum.

## Tables and Figures



**Figure 1. A model of writing proposed by Hayes and Flower (1980) and Hayes (1996): Planning, text generation, and revision (used by permission).**

**Table 1**  
**Data Description, Categorical Variables**

<b>Group</b>	<b>Number of Cases</b>	<b>Percentage</b>
Total	114	100
<b>Gender (Q1)</b>		
Female	89	78.1
Male	25	21.9
<b>Ethnic (Q2)</b>		
Black	6	5.3
Asian	6	5.3
White	96	84.2
Hispanic	4	3.5
Other	2	1.8
<b>English Dominant (Q3)</b>		
Yes	107	93.9
No	7	6.1
<b>English Better (Q4)</b>		
Yes	109	95.6
No	5	4.4
<b>Status (Q10)</b>		
Freshman	1	0.9
Sophomore	47	41.2
Junior	37	32.5
Senior	29	25.4
<b>Graduate School (Q12)</b>		
Yes	91	80.5
No	22	19.5

**Note.** The questionnaire items from which the data were derived are given in parentheses in the first column (Q1, Q2, etc.).

**Table 2**  
**Data Description, Continuous Variables**

Variable	N	Mean	S.D.	Minimum	Maximum
<b>Questionnaire variables</b>					
WABIL (Q5)	114	3.78	.70	2.00	5.00
AREA (Q6)	114	3.80	.49	2.50	5.00
GENRE (Q7)	114	3.83	.44	2.40	5.00
MATHGR (Q8)	114	5.18	1.30	1.00	7.00
ENGGR (Q8)	114	5.70	.92	3.00	7.00
RECGR (Q8)	114	6.02	.99	3.00	7.00
WACT (Q9)	114	7.97	4.79	0.00	22.00
GPA (Q11)	113	3.21	.42	2.07	4.00
SAT V score (Q13)	99	604.9	73.73	410	800
SATM score (Q13)	99	598.3	81.19	250	760
SAT W score (Q13)	24	642.9	87.05	400	800
<b>Editing task scores</b>					
First task score	112	17.56	5.52	4.00	26.00
Second task score	111	17.48	5.86	1.00	30.00
Total score (EDIT)	111	35.02	10.07	5.00	56.00
<b>Editing task times (Min.)</b>					
First task time	112	14.14	4.79	1.15	27.23
Second task time	111	11.98	4.66	3.57	29.44
Total time	111	26.06	8.38	7.03	51.65

**Variable Descriptions:**

EDIT = Total editing score

WABIL = Student self-rating of their writing ability as compared to peers (Questionnaire Item 5)

AREA = Student self-rating of college success in specific areas of writing (Questionnaire Item 6, average of responses)

GENRE = Student self-rating of college success in different writing genres (Questionnaire Item 7, average of responses)

ENGGR = English, history, and similar grades (Questionnaire Item 8)

MATHGR = Math, science, and similar grades (Questionnaire Item 8)

RECGR = Most recent writing grade (Questionnaire Item 8)

WACT = Writing activities and accomplishments (Questionnaire Item 9, weighted count)

GPA = College GPA (Questionnaire Item 11)

**Table 3**  
**Correlation Matrix, Excluding SAT Scores**

	EDIT	WABIL	AREA	GENRE	ENGGR	MATHGR	RECGR	WACT	GPA
EDIT	1.00								
WABIL	0.35	1.00							
AREA	0.28	0.66	1.00						
GENRE	0.20	0.43	0.67	1.00					
ENGGR	0.46	0.52	0.50	0.36	1.00				
MATHGR	0.31	0.20	0.26	0.19	0.31	1.00			
RECGR	0.29	0.41	0.42	0.28	0.58	0.15	1.00		
WACT	0.16	0.22	0.30	0.23	0.19	0.02	0.02	1.00	
GPA	0.35	0.33	0.36	0.19	0.52	0.68	0.42	0.10	1.00

**Note.** N = 102

**Variable Descriptions:**

EDIT = Total editing score

WABIL = Student self-rating of their writing ability as compared to peers (Questionnaire Item 5)

AREA = Student self-rating of college success in specific areas of writing (Questionnaire Item 6, average of responses)

GENRE = Student self-rating of college success in different writing genres (Questionnaire Item 7, average of responses)

ENGGR = English, history, and similar grades (Questionnaire Item 8)

MATHGR = Math, science, and similar grades (Questionnaire Item 8)

RECGR = Most recent writing grade (Questionnaire Item 8)

WACT = Writing activities and accomplishments (Questionnaire Item 9, weighted count)

GPA = College GPA (Questionnaire Item 11)

**Table 4**  
**Correlation Matrix, Including SAT Scores**

	EDIT	WABIL	AREA	GENRE	ENGGR	MATHGR	RECGR	WACT	GPA	SATV	SATM
<b>EDIT</b>	1.00										
<b>WABIL</b>	0.37	1.00									
<b>AREA</b>	0.26	0.70	1.00								
<b>GENRE</b>	0.14	0.44	0.61	1.00							
<b>ENGGR</b>	0.44	0.56	0.52	0.32	1.00						
<b>MATHGR</b>	0.25	0.16	0.24	0.18	0.29	1.00					
<b>RECGR</b>	0.28	0.40	0.39	0.27	0.56	0.17	1.00				
<b>WACT</b>	0.25	0.28	0.38	0.29	0.28	0.08	0.02	1.00			
<b>GPA</b>	0.32	0.33	0.33	0.12	0.54	0.67	0.44	0.15	1.00		
<b>SATV</b>	0.39	0.29	0.29	0.26	0.44	0.14	0.26	0.14	0.31	1.00	
<b>SATM</b>	0.36	0.24	0.14	0.20	0.32	0.29	0.20	0.02	0.25	0.41	1.00

**Note.** N = 93

**Variable Descriptions:**

EDIT = Total editing score

WABIL = Student self-rating of their writing ability as compared to peers (Questionnaire Item 5)

AREA = Student self-rating of college success in specific areas of writing (Questionnaire Item 6, average of responses)

GENRE = Student self-rating of college success in different writing genres (Questionnaire Item 7, average of responses)

ENGGR = English, history, and similar grades (Questionnaire Item 8)

MATHGR = Math, science, and similar grades (Questionnaire Item 8)

RECGR = Most recent writing grade (Questionnaire Item 8)

WACT = Writing activities and accomplishments (Questionnaire Item 9, weighted count)

GPA = College GPA (Questionnaire Item 11)

SATV = SAT I verbal score

SATM = SAT I mathematics score

**Table 5**

**Correlations Between First and Second Editing Task Scores and Other Variables**

<b>Variable</b>	<b>N</b>	<b>Correlation with first task</b>	<b>Correlation with second task</b>
WABIL	102	.37**	.29**
AREA	102	.34**	.19
GENRE	102	.24*	.14
ENGGR	102	.41**	.45**
MATHGR	102	.26**	.31**
RECGR	102	.24*	.29**
ACCOMP	102	.21*	.09
GPA	102	.29**	.34**
SATV	93	.34**	.36**
SATM	93	.35**	.32**
<b>Average correlation</b>		<b>.30</b>	<b>.28</b>

\*p < .05

\*\*p < .01

**Variable Descriptions:**

EDIT = Total editing score

WABIL = Student self-rating of their writing ability as compared to peers (Questionnaire Item 5)

AREA = Student self-rating of college success in specific areas of writing (Questionnaire Item 6, average of responses)

GENRE = Student self-rating of college success in different writing genres (Questionnaire Item 7, average of responses)

ENGGR = English, history, and similar grades (Questionnaire Item 8)

MATHGR = Math, science, and similar grades (Questionnaire Item 8)

RECGR = Most recent writing grade (Questionnaire Item 8)

WACT = Writing activities and accomplishments (Questionnaire Item 9, weighted count)

GPA = College GPA (Questionnaire Item 11)

SATV = SAT I verbal score

SATM = SAT I mathematics score

**Table 6**

**Item Analysis of Phase I and Phase II Editing Tasks**

Item	Task 1 (free speech)				Task 2 (materialistic desires)			
	Percent correct		Biserial correlation		Percent correct		Biserial correlation	
	Phase I	Phase II	Phase I	Phase II	Phase I	Phase II	Phase I	Phase II
1	.52	.17	-.01	.33	.17	.20	.35	.56
2	.60	.44	.21	.50	.44	.35	.42	.44
3	.81	.29	.43	.39	.29	.28	.36	.25
4	.25	.12	.46	.38	.12	.37	.32	.45
5	.52	.67	.11	.38	.67	.63	.46	.51
6	.71	.77	.22	.26	.77	.57	.27	.43
7	.02	.06	.25	.32	.06	.33	.18	.60
8	.81	.36	.26	.27	.36	.43	.37	.44
9	.67	.83	.53	.46	.83	.33	.27	.48
10	.90	.50	.19	.47	.50	.45	.44	.38
11	.36	.94	.54	.43	.94	.67	-.08	.48
12	.98	.44	.11	.14	.44	.37	.57	.48
13	.75	.98	.39	.23	.98	.92	.04	.14
14	.60	.36	.13	.42	.36	.43	.41	.25
15	.75	.36	.28	.11	.36	.28	.34	.58
16	.19	.90	.10	.50	.90	.18	.40	.12
17	.44	.65	.31	.62	.65	.57	.34	.22
18	.65	.50	.43	.33	.50	.43	.50	.38
19	.36	.54	.56	.38	.54	.69	.52	.43
20	.75	.81	.28	.31	.81	.61	.52	.53
21	.52	.56	.46	.45	.56	.55	.38	.39
22	.94	.64	.15	.34	.64	.51	.05	.41
23	.60	.65	.44	.44	.65	.03	.40	.04
24	.77	.27	.54	.25	.27	.67	.27	.32
25	.94	.19	-.02	.43	.19	.31	.04	.57
26	.48	.94	.29	.35	.94	.28	.10	.27
27	.73	.04	.39	.52	.04	.92	.37	.11
28	.44	.77	.41	.39	.77	.22	.30	.12
29	.21	.12	.31	.42	.12	.08	.46	.25
30	.35	.75	.42	.58	.75	.78	.39	.51
31	.44	.10	.34	.59	.10	.18	.34	.68
32	.21	.71	.29	.60	.71	.75	.46	.01
<b>Avg.</b>	<b>.57</b>	<b>.53</b>	<b>.31</b>	<b>.39</b>	<b>.53</b>	<b>.44</b>	<b>.33</b>	<b>.37</b>



**Table 7**  
**Taxonomy Analysis Results**

<b>Problem category</b>	<b>Problem</b>	<b>Number</b>
<b>Lexical problems</b>	Wordiness	20
	Redundancy	5
	Diction	6
	Weak phrase	2
	Lack of precision	2
	Inappropriate preposition	2
	<b>Category total</b>	<b>37</b>
<b>Grammar problems</b>	Plural possessive	2
	Relative pronoun (which/that)	2
	Homonym (there/their/they're)	1
	<b>Category total</b>	<b>5</b>
<b>Syntax</b>	Subject-verb agreement	4
	Verb tense	4
	Verb form	4
	Pronoun-antecedent agreement	2
	Subject of clause after participial phrase	1
	Subordination	1
	Active/passive voice	1
	Appositive	1
	<b>Category total</b>	<b>18</b>
<b>Punctuation problems</b>	Punctuation <i>(1 Discourse-oriented – Free Speech 4)</i>	6
	Sentence boundary <i>(1 Discourse-oriented – Free Speech 23)</i>	2
	<b>Category total</b>	<b>8</b>
<b>Coherence problems</b>	Logical transition <i>(6 Discourse-oriented – Free Speech 8,10, 13, Material Desires 7, 23, 25)</i>	8
	Logical connective <i>(1 Discourse-oriented – Material Desires 2)</i>	3
	Logical relationship (cause/effect)	1
	Parallelism	2
	Correlative conjunction (non only/but also)	1
	<b>Category total</b>	<b>15</b>
	<b>Grand total</b>	

## **Appendix A**

### **Editing Tasks, Student Instructions, and Scoring Procedures**



## Student Instructions

(You will be able to review these instructions during the test.)

The following passages contain a number of bracketed sections, each of which may or may not include composition errors or sections that can be improved. These errors may include errors in grammar, punctuation, syntax, sentence structure, vocabulary, idiom, or word choice. Sections may also contain unnecessary or redundant words or phrases, run-on sentences, sentence fragments, or poor transitions from the preceding sentence or paragraph. Bracketed sections may contain no errors, however, and may not need improvement.

If the bracketed section in the passage contains an error, or if you believe it can be improved, click on any word in the phrase. The word or phrase will appear in the edit window at the bottom of the screen. You can then change the word or phrase using simple word processing techniques (back-space, delete, and type-in). When you have made any change you think appropriate, click on the "**Replace Phrase**" button located to the right of the edit window. The changed word or phrase, along with associated punctuation, will replace the original word or phrase in the passage. After you have changed a bracketed section, use the cursor to go on to the next bracketed section. If you believe that a bracketed section contains no errors or cannot be improved, simply ignore it and go to the next section.

If you wish to go back or forward to change a bracketed section you have already corrected, simply use the cursor to return to that section. If you would like to see the original passage before your corrections, click on the "**View Original Passage**" button. To revert to the passage you are correcting, click on the "**Return to Editing**" button on your screen.

When you have finished correcting all the bracketed sections, click on the "**Next**" icon on the bottom right of your screen.

**Important:** When making changes, the following rules apply:

1. Do not change the basic meaning of a sentence.
2. Use as few words as possible to correct errors or to improve the writing while preserving its original meaning.
3. Do not alter a bracketed section unless you think you can improve it.
4. You may not make changes other than within the bracketed sections.

## Task One: Free Speech

The right to free speech, (1) **[which is a]** founding principle of the United States, was made law by the signers of the Constitution. (2) **[It is my opinion that we]** must protect this fundamental right at any cost. Some claim that order and civility (3) **[requires]** limits upon free (4) **[speech; the]** danger of restricting our open expression far outweighs the possible benefit (5) **[to society generally]**.

America (6) **[is]** founded upon the entitlement (7) **[to]** life, liberty, and the pursuit of (8) **[happiness: but]** can a citizen possess liberty when (9) **[we lack]** the power to voice opinions? (10) **[Although a]** strong government must be willing to risk the dissent of (11) **[there]** people. Failure to safeguard free expression threatens the freedom of the (12) **[individual, and]** also the health of the society.

(13) **[Therefore,]** maintaining a balance between the individual right to free speech and the rights of the group presents government with a dilemma. The (14) **[problem and]** complication of having free speech is that some opinions can be controversial and even offensive to other citizens. Speech (15) **[would have]** the power (16) **[as inspiration]** and to (17) **[be injurious]**. But wounding (18) **[others]** sensibilities (19) **[would not be]** the same as harming (20) **[others']** physically. Our culture must treat rude or controversial ideas with constructive criticism rather than (21) **[fixing them with bans]**. After all, ideas exist (22) **[irregardless]** who has (23) **[them and]** we cannot refuse to hear them. Restricting free speech would only (24) **[give encouragement to]** inflammatory notions to go (25) **[underground, and so depriving]** society of the chance to deal with them openly.

History (26) **[definitely shows]** the importance of free speech in a healthy society. Rulers need to be criticized when they behave unfairly or (27) **[in an unwise way]**. Corrupt regimes (28) **[which]** impose limits on free speech (29) **[for remaining]** in power should be removed. (30) **[In the final analysis, vigorous]** debate helps us to grasp opposing views (31) **[and it helps us to understand]** more clearly our own (32) **[position on any given question]**.

### One corrected version:

The right to free speech, a founding principle of the United States, was made law by the signers of the Constitution. We must protect this fundamental right at any cost. Some claim that order and civility require limits upon free speech, but the danger of restricting our open expression far outweighs the possible benefit to society.

America was founded upon the entitlement to life, liberty, and the pursuit of happiness, but how can a citizen possess liberty when he or she lacks the power to voice opinions? A strong government must be willing to risk the dissent of its people. Failure to safeguard free expression threatens the freedom of the individual and also the health of the society.

Nevertheless, maintaining a balance between the individual right to free speech and the rights of the group presents government with a dilemma. The complication of having free speech is that some opinions can be controversial and even offensive to other citizens. Speech has the power to inspire and to injure. But wounding others' sensibilities is not the same as harming them physically. Our culture must treat rude or controversial ideas with constructive criticism rather than bans. After all, ideas exist regardless of who has them, and we cannot refuse to hear them. Restricting free speech would only encourage inflammatory notions to go underground, depriving society of the chance to deal with them openly.

History shows the importance of free speech in a healthy society. Rulers need to be criticized when they behave unfairly or unwisely. Corrupt regimes that impose limits on free speech to remain in power should be removed. Vigorous debate helps us to grasp opposing views and to understand more clearly our own position.

### Sample Scoring Key

[Phrase9Key]

NoKeys=14

Key1=he or she lacks

Key2=he/she lacks

Key3=she or he lacks

Key4=she/he lacks

Key5=a citizen lacks

Key6=s/he lacks

Key7=h/she lacks

Key8=that person lacks

Key9=that citizen lacks

Key10=one lacks

Key11=lacking

Key12=he lacks

Key13=she lacks

Key14=society lacks

## Task Two: Materialistic Desires

(1) **[Despite the fact that]** it is only human to have materialistic desires, some believe this to be wrong. It is only natural to want to improve one's material (2) **[circumstances and]** it is the (3) **[consensus of opinion]** in capitalistic societies that there is nothing wrong with it. Human nature desires both progress and (4) **[recognition, material]** goods are tangible symbols of hard work. The desire for things (5) **[cannot but help]** the individual by (6) **[it's encouragement of]** habits of hard work and perseverance. (7) **[However]** materialism benefits the culture as well.

If (8) **[it were the case that]** people were always contented with their present (9) **[ways]** and never

(10) **[gave consideration to]** a better lifestyle, life would be very dull. New inventions (11) **[resulting from]** the urge to make life more comfortable and interesting and to make ourselves (12) **[more attractive in appearance]**. Materialistic desires (13) **[make]** us to use our brains and (14) to **[imagine]** new tools that ease (15) **[the burdens of our days]**.

Not only do these desires spur invention, (16) **[however]** they also encourage exchange and trade (17) **[amongst nations]**. The quest for an easier life (18) **[by means of]** pursuing the "latest model" requires resourcefulness. If we (19) **[were]** to meet the demands of (20) **[the modern world of today]**, we must be clever. Goods (21) **[which]** we cannot find in our own country (22) **[are necessary to acquire]** through international trade. (23) **[Diverse]** societies learn about each other (24) **[owing to]** their materialistic desires.

(25) **[Nevertheless, the]** health of the economy depends upon our desire for material objects. By keeping up our demand for more (26) **[of a variety of]** products, (27) **[it can]** create jobs for (28) **[quite a few]** people. None of these jobs (29) **[are]** unnecessary (30) **[for the reason that]** they are essential to make the economy function. When (31) **[products are purchased]**, we help not only (32) **[ourselves but]** we also help others by providing them with jobs.

### One corrected version

Although it is only human to have materialistic desires, some believe this to be wrong. It is only natural to want to improve one's material circumstances, and it is the consensus in capitalistic societies that there is nothing wrong with it. Human nature desires both progress and recognition, and material goods are tangible symbols of hard work. The desire for things can only help the individual by encouraging habits of hard work and perseverance. Moreover, materialism benefits the culture as well.

If people were always contented with their present situation and never considered a better lifestyle, life would be very dull. New inventions result from the urge to make life more comfortable and interesting and to make ourselves more attractive. Materialistic desires spur us to use our brains and to invent new tools that ease our burdens.

Not only do these desires spur invention, they also encourage exchange and trade among nations. The quest for an easier life by pursuing the "latest model" requires resourcefulness. If we are to meet the demands of the modern world, we must be clever. Goods that we cannot find in our own country can be acquired through international trade. Thus, diverse societies learn about each other through their materialistic desires.

More importantly, the health of the economy depends upon our desire for material objects. By keeping up our demand for more varied products, we create jobs for many people. None of these jobs is unnecessary since they are essential to make the economy function. When we purchase products, we help not only ourselves, but we also help others by providing them with jobs.

### Sample Scoring Key

[Phrase1Key]

NoKeys=15

Key1=Although

Key2=Even though

Key3=Though

Key4=While

Key5=Even if

Key6=While most agree that

Key7=While many agree that

Key8=Albeit that

Key9=Even supposing

Key10=Even supposing that

Key11=Even though many think

Key12=Despite popular belief, that

Key13=Despite that

Key14=While most people believe that

Key15=Despite the strong belief that



## Scoring Procedures

Scoring of the editing tasks is accomplished by simple pattern matching of examinee responses with solutions stored in a database. The solutions are developed by means of a three-step process:

1. The experts who create the editing tasks list the solutions that occur to them.
2. Pretesting of the editing tasks reveals solutions that did not occur to the experts when the tasks were created. These solutions are added to the database following expert review.
3. Each time a task is administered, all nonmatching solutions are checked to determine if they are acceptable to experts. Any solutions that are found to be acceptable are added to the database.

Additional solutions can be generated by using a thesaurus to identify synonyms. These additional solutions need to be checked by a writing expert, however, before they are added to the list of keys. Only a limited number of thesaurus-generated solutions has been implemented to date, but automated language processing has significant potential for this purpose.

An example of an item from one of the two editing tasks (Materialistic Desires) illustrates how additional solutions can be generated using a thesaurus:

"(3)[consensus of opinion]"

[Phrase3Key]

NoKeys=49

Key1=consensus

Key2=general opinion

Key3=general view (thesaurus)

Key4=general viewpoint (thesaurus)

Key5=general belief (thesaurus)

Key6=general notion (thesaurus)

Key7=prevailing opinion (thesaurus)

Key8=prevailing view (thesaurus)

Key9=prevailing viewpoint (thesaurus)

Key10=prevailing belief (thesaurus)

Key11=prevailing notion (thesaurus)

Key12=widespread opinion (thesaurus)

Key13=widespread view (thesaurus)

Key14=widespread viewpoint (thesaurus)

Key15=widespread belief (thesaurus)

Key16=widespread notion (thesaurus)

Key17=dominant opinion (thesaurus)  
Key18=dominant view (thesaurus)  
Key19=dominant viewpoint (thesaurus)  
Key20=dominant belief (thesaurus)  
Key21=dominant notion (thesaurus)  
Key22=predominant opinion (thesaurus)  
Key23=predominant view (thesaurus)  
Key24=predominant viewpoint (thesaurus)  
Key25=predominant belief (thesaurus)  
Key26=predominant notion (thesaurus)  
Key27=popular opinion  
Key28=popular view (thesaurus)  
Key29=popular viewpoint (thesaurus)  
Key30=popular belief (thesaurus)  
Key31=popular notion (thesaurus)  
Key32=accepted opinion (thesaurus)  
Key33=accepted view (thesaurus)  
Key34=accepted viewpoint (thesaurus)  
Key35=accepted notion (thesaurus)  
Key36=majority opinion  
Key37=majority view (thesaurus)  
Key38=majority viewpoint (thesaurus)  
Key39=general consensus  
Key40=overall opinion  
Key41=majority's consensus  
Key42=belief (thesaurus)  
Key43=common opinion  
Key44=common view (thesaurus)  
Key45=common viewpoint (thesaurus)  
Key46=common belief (thesaurus)  
Key47=common notion (thesaurus)  
Key48=opinion of many  
Key49=common consensus



## **Appendix B**

### **Writing Experience Questionnaire**



## Writing Experience Questionnaire

1. What is your gender?
  - a. Female
  - b. Male
  
2. How do you describe yourself?
  - a. African American/Black (non-Hispanic)
  - b. American Indian/Native American/Alaskan Native
  - c. Asian American/Pacific Island American
  - d. Caucasian/White (non-Hispanic)
  - e. Hispanic/Latino/Chicano/Mexican American/Puerto Rican
  - f. Other
  
3. Is English the dominant language in your household?
  - a. Yes
  - b. No
  
4. Do you understand English as well as or better than any other language?
  - a. Yes
  - b. No
  
5. How well do you think your English composition skills compare with those of other college students?
 

Well above average	..... 5
Somewhat above average	..... 4
About average	..... 3
Somewhat below average	..... 2
Well below average	..... 1

6. For each of the activities listed, please indicate how successful you have been in college writing assignments. (Circle one number for each.)

5 = extremely successful      3 = somewhat successful      1 = not at all successful  
 4 = quite successful      2 = not very successful

A. Thinking about an assignment (e.g., developing ideas, gathering information)	5	4	3	2	1
B. Organizing (e.g., making outlines, deciding on the order to present ideas)	5	4	3	2	1
C. Revising (e.g., improving sentence phrasing, rearranging ideas, correcting grammar and punctuation)	5	4	3	2	1
D. Developing an effective writing style (e.g., expressing ideas clearly and in interesting ways)	5	4	3	2	1

7. For each of the kinds of writing listed, please indicate how successful you have been in college.

5 = extremely successful      3 = somewhat successful      1 = not successful at all  
 4 = quite successful      2 = not very successful      0 = I did not do this

A. Personal writing (e.g., describing experiences or feelings as in a journal)	5	4	3	2	1	0
B. Creative writing ( e.g., a poem or short story)	5	4	3	2	1	0
C. Persuasion (e.g., arguing a position or writing a letter to the editor)	5	4	3	2	1	0
D. Analysis/criticism (e.g., reviewing a book, movie, article, theory, work of art)	5	4	3	2	1	0
E. Description/explanation (e.g., describing an experiment or how something works)	5	4	3	2	1	0
F. Examination writing (e.g., long essay answers)	5	4	3	2	1	0
G. Applied writing (e.g., preparing an application or resume)	5	4	3	2	1	0

8. Approximately, what grades have you received in college: (A) in math, science, or similar courses not requiring much writing, (B) in English, history, and similar courses often requiring much writing, (C) on the most recent assignment for which you had to write a report, essay, etc. If your course did not use letter grades, please mark the grade that is the closest equivalent. (Circle one number for each.)

Grade	(A) Math, science, and similar courses	(B) English, history, and similar courses	(C) Most recent writing assignment
A . . . . .	7	7	7
A- . . . . .	6	6	6
B . . . . .	5	5	5
B- . . . . .	4	4	4
C . . . . .	3	3	3
C- . . . . .	2	2	2
D or less . .	1	1	1

9. Writing Activities

Descriptions of a variety of writing activities in school or college, out of school or college, in volunteer work, or in part-time or full-time jobs are listed below. Please read each description, and then indicate whether you have engaged in the activity by checking the "Yes" or "No" box next to the description. Some of the activities are relatively uncommon ones in which few people have been engaged.

- |                          |                          |   |
|--------------------------|--------------------------|---|
| Yes                      | No                       |   |
| <input type="checkbox"/> | <input type="checkbox"/> | 1. Had a pen-pal for less than a year.  |
| <input type="checkbox"/> | <input type="checkbox"/> | 2. Had a pen-pal for more than a year.  |
| <input type="checkbox"/> | <input type="checkbox"/> | 3. Kept a nonschool journal or a diary for less than a year.                    |
| <input type="checkbox"/> | <input type="checkbox"/> | 4. Kept a nonschool journal or a diary for more than a year.                    |
| <input type="checkbox"/> | <input type="checkbox"/> | 5. Wrote a letter to the editor of a student newspaper that was published.      |
| <input type="checkbox"/> | <input type="checkbox"/> | 6. Wrote a letter to the editor of a town or city newspaper that was published. |
| <input type="checkbox"/> | <input type="checkbox"/> | 7. Wrote a letter to the editor of a magazine that was published.               |
| <input type="checkbox"/> | <input type="checkbox"/> | 8. Worked as a reporter for a student newspaper.                                |
| <input type="checkbox"/> | <input type="checkbox"/> | 9. Worked on the editorial staff of a student newspaper.                        |



Question 9 (continued):

- | Yes | No  |   |
|-----|-----|---|
| [ ] | [ ] | 10. Served as an assistant or associate editor of a student newspaper.    |
| [ ] | [ ] | 11. Served as the editor for a student newspaper.                         |
| [ ] | [ ] | 12. Worked on the editorial staff for a student yearbook.                 |
| [ ] | [ ] | 13. Served as an assistant or associate editor for a student yearbook.    |
| [ ] | [ ] | 14. Served as the editor for a student yearbook.                          |
| [ ] | [ ] | 15. Worked as a reporter for a town or city newspaper.                    |
| [ ] | [ ] | 16. Participated in an essay contest.                                     |
| [ ] | [ ] | 17. Won an essay contest.   |
| [ ] | [ ] | 18. Wrote a feature article, column, or editorial that was published.     |
| [ ] | [ ] | 19. Wrote a speech that you gave at a public gathering.                   |
| [ ] | [ ] | 20. Wrote a speech for someone else that was given at a public gathering. |
| [ ] | [ ] | 21. Wrote advertising or public relations material.                       |
| [ ] | [ ] | 22. Wrote technical manuals or other instructional material.              |
| [ ] | [ ] | 23. Wrote poetry, fiction, or essays that were published.                 |
| [ ] | [ ] | 24. Wrote a play that was publicly performed.                             |
| [ ] | [ ] | 25. Wrote a screenplay for a film.  |
| [ ] | [ ] | 26. Wrote a script for a dramatic or comedy show.                         |
| [ ] | [ ] | 27. Attended a writer's workshop.   |
| [ ] | [ ] | 28. Wrote a short story that was not a part of school work.               |
| [ ] | [ ] | 29. Wrote a novel or other full-length book.                              |
| [ ] | [ ] | 30. Other writing experience or activity: _____                           |

10. What is your status in college: (a) Freshman (b) Sophomore (c) Junior (d) Senior
11. What is your college GPA on a 0 to 4 scale (for example, 3.35)? \_\_\_\_\_.
12. Do you think you will go to graduate school? (a) Yes (b) No
13. What were your SAT scores? Verbal \_\_\_\_\_ Math \_\_\_\_\_ Writing \_\_\_\_\_



## **Appendix C**

### **Outline of the Preliminary Taxonomy of Writing Principles and Flaws**



1. Lexical Problems
  - 1.1 Sentence-Oriented Lexical Problems
    - 1.1.1 Words and Phrases
      - 1.1.1.1 Specific Inappropriate Word or Phrase
      - 1.1.1.2 General Word Usage
        - 1.1.1.2.1 Redundancy
        - 1.1.1.2.2 Wordiness
        - 1.1.1.2.3 Diction
        - 1.1.1.2.4 Weak phrase
        - 1.1.1.2.5 Lack of Precision
    - 1.1.2 Idioms & Cliches
      - 1.1.2.1 Specific Inappropriate Idioms and Cliches
      - 1.1.2.2 General Idioms and Cliches
    - 1.1.3 Spelling
  - 1.2 Discourse-Oriented Lexical Problems
    - 1.2.1 General Lexical Items
    - 1.2.2 Idioms and Cliches
2. Grammar Problems
  - 2.1 Sentence-Oriented Grammar Problems
    - 2.1.1 Plural Possessives and Apostrophes
    - 2.1.2 Relative Pronouns (which/that)
    - 2.1.3 Homonyms (there/their/they're)
    - 2.1.4 Parts of Speech
    - 2.1.5 General
  - 2.2 Discourse-Oriented Grammar Problems
    - 2.2.1 Action-Object Expression
3. Syntax Problems
  - 3.1 Sentence-Oriented Syntax Problems
    - 3.1.1 Subject-Verb Agreement
    - 3.1.2 Verb Tense and Verb Form
    - 3.1.3 Pronoun-Antecedent Agreement
    - 3.1.4 Active/Passive Voice
    - 3.1.5 Subordination
    - 3.1.6 One-point-per-clause
    - 3.1.7 Parallelism
  - 3.2 Discourse-Oriented Syntax Problems
    - 3.2.1 Verb Tense and Verb Form
    - 3.2.2 Pronoun-Antecedent Agreement
    - 3.2.3 Parallelism
4. Punctuation Problems
  - 4.1 Sentence-Oriented Punctuation Problems

- 4.1.1 Sentence Boundary
- 4.1.2 Comma Use
- 4.1.3 General
- 4.2 Discourse-Oriented Punctuation Problems
  - 4.2.1 Sentence Boundary
  - 4.2.2 Comma Use
  - 4.2.3 Semicolons
  - 4.2.4 Clauses & Conjunctions
  - 4.2.5 General
- 5. Coherence Problems
  - 5.1 Sentence-Oriented Coherence Problems
    - 5.1.1 Logical Transitions
    - 5.1.2 Logical Connectives
    - 5.1.3 Logical Organization
    - 5.1.4 Parallelism
    - 5.1.5 Conciseness
    - 5.1.6 General
  - 5.2 Discourse-Oriented Coherence Problems
    - 5.2.1 Logical Transitions
    - 5.2.2 Logical Connectives
    - 5.2.3 Logical Organization
    - 5.2.4 Parallelism
    - 5.2.5 Conciseness
    - 5.2.6 Verb Tense And Verb Form
    - 5.2.7 Reader Expectations
    - 5.2.8 General
- 6. Style
  - 6.1 Discourse-Oriented Style Problems

