This investigation sought normative longitudinal change in student writing during college. It used a random sample of students (N = 64), each of whom had produced essays at two points in their undergraduate careers, matriculation and junior year. Measures were writing features showing undergraduate change toward competent, working-world performance. From a principal-components factoring of variables used in a previous study, nine measures were selected as good representatives of nine factors—factors of independent and bound ideas, idea elaboration and substantiation, local cohesion, establishment of logical boundaries, free modification, fluency, and vocabulary. When applied to the 1st-year and junior-year writing, eight of the nine measures, including a holistic rating, recorded statistically significant change, all in the direction of workplace performance. Directions for further research are discussed.

Documenting Improvement in College Writing

A Longitudinal Approach

RICHARD H. HASWELL
Texas A&M University–Corpus Christi

CHANGE IN COLLEGE STUDENT WRITING: SOME LONGITUDINAL FINDINGS

Three and a half decades ago, Albert Kitzhaber (1963) launched an unofficial project to document the ways that college students in the United States change in their writing performance during the undergraduate years. Since then, it is only fair to say the project has not been much advanced by researchers. Findings have been more critiqued than replicated or advanced. New methods of investigation have

Author's Note: This study was supported in part by the General Education Office at Washington State University, and by the Mary and Paul Haas Endowment to the English Program at Texas A&M University–Corpus Christi. For statistical assistance, thanks to Dr. James Whipple, Department of Psychology, Washington State University, and to Dr. Edward R. Jones, Director of the Center for Statistical and Quality Improvement Services at Texas A&M University–Corpus Christi.

WRITTEN COMMUNICATION, Vol. 17 No. 3, July 2000 307-352

tended to abandon old methods rather than assimilate them. Even scholars committed to serious research into college writing have expressed discouragement with the project. For instance, in a summary of efforts to document the effect of instruction on student writing—one of Kitzhaber's initial goals—White (1989) asserts that "there is no replicated design in existence for demonstrating that any writing instructional program in fact improves student writing, if we define writing in a sophisticated way" (p. 198).

To this discouragement, the current article offers a modest counter. It reports a study of writing that uses quantified measures, most of them drawn from previous research, and that records statistically significant changes in the unrehearsed writing of university students from matriculation to early in their junior year. The study makes no claims about effects of instruction on these changes in student performance. But the measures used by the study have been associated with the performance of competent postgraduate writers in business and industry, so, to that extent, the findings help qualify the direction that the undergraduate writing analyzed seems to be moving. In short, the study adds a positive piece to the investigation of change in undergraduate writing. In its application of previously tried measures combined with a new design, it also suggests that past research still has its uses. This article ends with some recommended lines of future research that might eventually turn Kitzhaber's (1963) project into something more promising than, as White pictures it, a "treacherous shoal" littered with foundered "evaluation barks" (p. 201).

THE RESEARCH DESIGN

Limitations of Past Research

In its design, this study was aware of some of the limitations of previous efforts to identify undergraduate change in writing. In the past, a major difficulty has been in finding ways to appraise student writing that are feasible, replicable, and meaningful. Analysis applying a large battery of measures, in the spirit of exploration, often ended up with results that were difficult to interpret because of undefined multicollinearity (Freedman & Pringle, 1980; Gebhard, 1978; Haswell, 1986; Kitzhaber, 1963). What looked like a panorama of writing skills actually might be quite patchy because clusters of measures were

highly correlated with one another. On the other hand, studies that focus on one measurable quality of writing have problems in relating the analysis to a generalized sense of writing competence: for example, Hagen (1971) (transitions), Hays (1982) (sense of audience), Hunt (1970) (t-unit and clause size), Maimon and Nodine (1979) (syntactic errors), Schumacher, Klare, Cronin, and Moses (1984) (pauses while composing). Stewart (1978) found words per t-unit increasing steadily through 1st-, 2nd-, 3rd-, and 4th-year undergraduates with a statistically significant jump beyond that with graduate students; but the stability of t-unit size has been little explored with research groups (Witte & Davis, 1982); nor has the relationship of t-unit length with logic, support, audience awareness, rhetorical mode, and other factors of successful writing (Faigley, 1980). The same is true with other measures of syntax and writing quality. A common finding, for instance, is that any single syntactic measure adds little to a prediction of holistic scores (e.g., Grobe, 1981; Witte & Faigley, 1981). Yet, even the holistic itself may not serve as well as was once thought as a single measure of writing change. Vehicle as it is for subjective impressions, holistic scoring may be insensitive to certain components of a developing writing skill (Haswell, 1991; Scharton, 1989) and probably hides its own strain of multicollinearity in the form of halo effects (Freedman, 1979; Haswell, 1999).

Another difficulty is establishing a convincing terminus ad quem for measures of change. At what point, for instance, does more and more free modification (Christensen & Christensen, 1967) or time spent paused in intermediate planning while composing (Schumacher et al., 1984) reach counterproductive levels? It only begs the questions to assume that the oldest student group under investigation stands as a terminus, because change over time cannot be assumed if it is to be proved. And to establish a group of skilled or professional writers as terminus, as much research has done, usually forces a comparison of performances under quite different circumstances and motives (e.g., Christensen & Christensen, 1967; Gebhard, 1978; Hagen, 1971; Haswell, 1986; Hunt, 1970).

The need to compare groups who may be different in uncompared ways is, of course, the Achilles heel of cross-sectional designs. The trouble is most obvious when researchers select only one undergraduate group and then compare it with a precollege group, perhaps high school seniors, or with a postcollege group, perhaps graduate students or professors, or even with both (Dixon, 1970; Evans, 1979; Evans & Ballance, 1980; Freedman & Pringle, 1980; Hunt, 1970;

Newberry, 1967; Watson, 1983). Any change discovered in writing performance may simply reflect self-selection into the group, especially in terms of the discursive skills needed to qualify as a member. And with only one college group sampled, normative changes in writing during the 4 undergraduate years can only be guessed. Even when research studies groups spread across the undergraduate years, it compares groups that are not cohorts, who again probably did not produce writing under very similar circumstances. Graham (1987), Kitzhaber (1963), Scharton (1989), and Schumacher et al. (1984), for instance, elicit writing from 1st-year and upper-division students in quite different courses.

Longitudinal designs are superior to cross-sectional in measuring group changes over time (Rest, 1979), but they have their own problems. When groups are studied, attrition of the original sample can be debilitating. Kerek, Daiker, and Morenberg (1980) lost 52% when they retested their 1st-year writers as juniors; Maimon and Nodine (1979) lost 43%. Wolcott (1994) started with 139 students, 1 year's special admissions class at a large university, and ended 3 years later with 6 students at graduation. Sternglass (1997) began with three 1st-year composition classes assigned to her at SUNY City College, found 53 students willing to engage in her study, and at graduation had 9. Generalization from such attenuated groups is hazardous. The longitudinal research published in the 1990s, largely ethnographic and largely suspending inquiry into group traits, suffers from its own handicap by confining its scrutiny to small numbers of participants. Chiseri-Strater (1991) followed two students for a couple of years, Haas (1994) one student for 4 years, Wolcott (1994) six students for 4 years, Spack (1997) one student for 3 years, Sternglass (1993, 1997) nine students for the 4 to 6 years it took all of them to graduate, and Herrington and Curtis (2000) three students for the 4 or 5 years it took them to graduate. Moreover, most of these participants occupy minority positions. Haas's student was a female biology major at Carnegie-Mellon who had graduated 14th in her high school class of 450. Chiseri-Strater (1991) "located two willing participants" out of an upper-division writing class (p. xxiii). Almost all of Wolcott's (1994), Spack's (1997), and Sternglass's (1993, 1997) students were classified as basic or second-language by their universities.

Finally, that these five qualitative studies do not share any measures of change is, unfortunately, characteristic of research into undergraduate writing change as a whole. Most of this ethnographic inquiry was open-ended interviewing or intuitive evaluation of

course writings, and the few empirical measures applied, such as Wolcott's (1994) use of the Nelson-Denny Reading Test or Haas's (1994) insightful quantified measures of authorship and writer agency, have not yet been compared across research venues. The same tends to be true with older studies. Only a few measures of writing product (e.g., essay size, t-unit and clause size, cohesive devices, free-modification use) have been much shared, as the present study will show. So far, composition research has not advantaged itself of the method of reapplying the same measure of human performance across many different contexts, although that method is a hallmark of most social science research.

Design of the Present Study

This critique of past research helped shape the current study. The design adds a longitudinal perspective to a series of investigations, conducted by the author, into undergraduate writing change. Originally, a cross-sectional design was used with more than a hundred measures of writing, reported in Haswell (1986). Haswell (1990) sets forth a critique of this design, including a factoring of the original battery of measures. Both study and critique are used to explore the connection between undergraduate writing and theories of personal development discussed in Haswell (1991). The current study uses the critique, particularly the factoring, to choose a reduced set of measures with a new sample of students. The main intent is to record longitudinal evidence of changes in their writing during the first 2½ years in college. The main hypothesis is that the cross-sectional evidence of undergraduate change toward postgraduate competent performance found in the earlier exploratory studies will be supported in the new longitudinal study.

Supported, it is hoped, with a tighter and inferentially more powerful design. To avoid the problems of a single measure or a battery of possibly redundant measures, it uses the factoring of the assortment of variables in Haswell (1986) and selects a short menu of nine measures. One of the nine is a holistic rating, but the other eight aid its interpretation because they are variables that the holistic probably does not much account for. To assist replication, all measures are empirical. To set a meaningful terminus, all the variables have a previous record of statistically significant undergraduate change toward the performance of competent, postgraduate workers writing at the

same task under similar conditions (Haswell, 1986). To enhance interpretation of outcomes, most of the nine measures have some record of shared use among researchers, with similar contrasts between undergraduate and skilled postgraduate writing recorded in a variety of research venues. To project group norms, a fairly sizable sample of undergraduates is randomly selected ($N = 64$). To avoid some of the problems of cross-sectional designs, the study is longitudinal, with writing performance elicited for each participant at two points in an undergraduate career—at entry into college and during the first semester of the junior year. To eliminate attrition of participants from the random sample, selection of participants takes place at the second point in time.

This report of the study proceeds in three steps. First is an analysis of the factoring of variables applied in the earlier study of undergraduate students (Haswell, 1986). Second is a selection of variables from the factoring to form a reduced menu of measures. Third is the application of the menu to the new population of undergraduates. No research design is perfect, of course, and some of the limitations of this one will be discussed, along with suggestions for ways research into undergraduate writing change might well continue.

STEP 1: A FACTORING

The previous cross-sectional study (Haswell, 1986) measured the writing of three groups of students, sampled at one point in time: entering college, entering 2nd year, and entering 3rd year. Students in the first group were 18 years of age; in the second, 19; and in the third, 20. All were students at Washington State University (WSU), a mid-sized, land-grant institution. The final comparison group was composed of college graduates who were 35 years old or older, working in business and industry, and judged by their supervisors as competent writers. All groups wrote impromptu on the same topics, for the same amount of time. To compare group performance, more than 100 variables were used, some devised during analysis. All were quantified measures, including holistic rating and counts of ideas, support, organization, diction, syntax, and mechanics.

To this empirical battery, a principal-component factor analysis was later applied (Haswell, 1990). The aim was to extract for use variables that were nonredundant. Factoring was also expected to aid

interpretation and selection of variables for subsequent studies by showing the relationship of underlying components to surface measures. All measures not preset before the original analysis in Haswell (1986) were excluded from the factoring. Of the 82 variables that remained (listed in Appendix A), three separate factorings were run. All three factorings used the principal-components computation of Harman (1976). They were not rotated because the data set did not meet Thurstone's (1947) criteria for simple structure.

The first run used 80 of the 82 variables. The 2 variables excluded were the holistic rating, on the chance that it was dependent on the other traits, and essay size in words, because number of words was the standard denominator used to compute rates for most of the other measures and consequently might generate a primary index effect. The second run used all 82 variables, reinstating the holistic and the measure of essay size, to identify components with which those two popular measures of writing might correlate highly. The third run used 74 variables, eliminating not only the 2 variables ejected from the first run but 6 others, all "basket" variables that collect the analysis of other measures under one sum, such as logical indicators, cohesive ties, and measures of free modification. The intent was to control for inflated whole/part effects. As it turned out, the factoring was essentially the same in all three runs, although differences in factor loadings did aid in subsequent analysis of the results. The results of the first run will be reported here, with allusions to the second and third runs where informative.

Kaiser-Guttman's rule recommends that interpretation of a factoring should consider only factors with Eigenvalues of 1.0 or better (Loehlin, 1992). The analysis extracted 23 such nontrivial factors, accounting for 93% of total variance. Of these 23 factors, 9 did not contain at least one variable with an acceptable loading, of .400 or better (Harman, 1976). The remaining 14 factors are listed in Table 1, each with its top-loading variable. Statistically, of course, these 14 factors are orthogonal to each other, and the top-loading variables together stand as best choice for a nonredundant cluster representing the full set of 82. The 14 factors account for 77% of total variance. Table 1 shows that this factoring provides a fairly strong support for an hypothesis of broad-based change in writing during and after the college years. Of the 14 factors, 9 show significant difference among groups, either among the three undergraduate groups or between one of the undergraduate groups and the postgraduates.

Table 1
Fourteen Factors Extracted From a Principal-Components Factoring of Measures of Writing

Factor	Eigenvalue	Variable (%)	Top-loading Variable	Loading
1	9.98	12.5	54. t-unit size[b]	.954
2	9.07	11.3	55. Clause size[b]	.831
3	6.69	8.4	08. Logical depth[a, b]	.712
4	5.01	6.3	61. Relative clauses[a]	.636
5	4.49	5.6	26. Logical indicators	.578
6	4.44	5.6	02. Rhetorical mode	–.608
7	4.10	5.1	71. Free modification[b]	.564
8	3.32	4.2	46. Passives	.571
9	3.08	3.9	06. Level 4 abstractions	.650
10	2.69	3.4	30. Synonym ties	.581
11	2.54	3.2	11. Exemplification[a]	.550
12	2.43	3.0	04. Size of logical pattern[a, b]	.479
13	2.28	2.9	85. Syntactic parallelism[a]	.415
18	1.44	1.8	95. Correct predication[b]	.404

NOTE: Numbering of variables indexes a full description of measures (Haswell, 1986).
a. Indicates that the measure was 1 of the 21 showing statistically significant differences between student groups (1st-, 2nd-, and 3rd-year students).
b. Indicates that the measure was 1 of the 44 showing a statistically significant difference between the student groups and postgraduate groups.

The top-loading variable, however, is of limited use in interpreting its factor. One variable, even with a very high loading, cannot divulge very clearly the quality of writing represented by the factor. Scrutiny of the other high-loading variables for each factor is needed to help define that factor. Such scrutiny is also needed for this study's particular goal of selecting a measure representative of each factor yet also with a good chance of recording hypothesized changes in undergraduate writing toward professional levels. The 14 factors will be taken up in order, with lists of variables loading at .400 or better, and with a brief characterization of the factor (an expanded analysis can be found in Haswell, 1990). Here and in subsequent lists, superscript *a* indicates that the variable showed statistically significant differences among student groups (1st-, 2nd-, and 3rd-year undergraduates), and superscript *b* indicates that the variable showed a statistically significant difference between the student groups and the postgraduates (Haswell, 1986). The variable number indexes the list in Appendix A. Decimal figures are factor loadings.

Factor 1: *Span of Multiclause Units*

54. t-unit size	.954[b]
53. Sentence size	.941[a, b]
66. Independent clauses	−.724
56. Clause/t-unit ratio	.721
65. Monoclauses	−.662
87. t-unit variance	.659[b]
46. Second-person plural	−.621[a, b]
86. Dependent sentence openers	.592
69. Postnominal modification	.537[a, b]
99. Conventional punctuation of main clauses	.514
70. Pre/postnominal modification ratio	.503[a, b]
68. Variance in nominal complexity	.502[a, b]
80. Prepositional strings	.476[b]
63. That clauses	.492
57. Coordination of t-units	−.483
71. Free modification	.481[b]
67. Nominal modification	.476[a, b]
60. Subordinate clauses	.455
76. Initial adverbs	.444
58. Predicate coordination	−.450
97. Conventional parallelism	.428[b]
42. Passives	.410

This factor, Factor 2, and Factor 7 partial out a mass of syntactic measures. Roughly, Factor 1 represents what might be called *multiclause-unit span*, the breadth of dependent ideas a writer is able or willing to encompass within a unit that the writer conceives of as syntactically independent. Sentence length and t-unit length stand in nearly perfect correlation with the factor, allowing easy interpretation. Percentage of total words put into independent clauses and into single-clause t-units express the factor negatively because these stylistic preferences co-occur with a bent for shorter independent syntactic units: the fewer the words put in dependent structures, the shorter the independent structures of sentence and t-unit tend to be. The high and positive loading of the variable of clause/t-unit ratio helps clarify the kind of syntactic breadth this factor represents, because that variable increases as more and more dependent clauses are attached to t-units. Rates of dependent sentence-openers, modifiers, prepositional phrases, free modifiers, adverbial phrases, and subordinate clauses all load positively because they form dependent structures that increase the size and complexity of t-units and sentences. The solid picture of group change with this factor supports Hunt's (1965, 1970) studies of the t-unit.

Factor 2: *Subclause Span of Bound Ideas*

55. Clause size	.831[b]
67. Nominal modification	.675[a, b]
65. Monoclauses	.622
56. Clause/t-unit ratio	−.600
79. Prepositions	.596[a, b]
70. Pre/postnominal modification ratio	.594[a, b]
38. Long words	.586[a, b]
86. Dependent sentence openers	−.568
60. Subordinate clauses	−.558
66. Independent clauses	.544
98. Standard sentences	.535
80. Prepositional strings	.534[b]
92. Conventional spelling	.512[b]
44. Pronominalization	−.507[b]
64. Base clauses	−.468
37. Infrequent words	.474[a, b]
58. Predicate coordination	−.467
10. Qualifiers	−.465
94. Conventional contractions	.424
76. Initial adverbs	−.432
99. Conventional punctuation of main clauses	.427
04. Top-level logical pattern	.416[a, b]
75. Adverbs	−.414[a]

Factor 2 represents subclause-unit span, the ability to conceive and utilize syntactically bound nodes of ideas. The factor supports Mellon's (1979) argument that one major feature of syntactic maturity is a growing attraction toward "restrictive structures of dominant noun phrases" (p. 17). Size of clause stands, of course, as the central syntactic measurement of this attraction, and bound nominal modification—as Mellon saw—the main syntactic tool to satisfy it. Generally, the other loadings with Factor 2 offer a forthright reading, with nominal modification and prepositional phrases (a major vehicle of bound modification) correlating positively, and dependent sentence openers, subordinate clauses, predicate coordination, qualifiers (usually nonrestrictive syntactically), and adverbial constructions correlating negatively. Factor 2 encompasses several important changes during the undergraduate years (in nominal modification, prepositions, vocabulary sophistication), and, compared with Factor 1, it marks an even more pervasive change in the postgraduate writing.

Factor 3: *Logical Elaboration of Ideas*

08. Logical depth	.712[a, b]
29. Identical cohesive ties	−.600
97. Conventional parallelism	.580[b]
40. Elegant variation	.571
13. Level 1 generalizations	.530[a]
07. Size of second-level logical pattern	.530[b]
28. Cohesive ties	−.509[b]
05. Size of top-level logical pattern	.475[b]
17. Unique events	.470
70. Pre/postnominal modification ratio	−.461[a, b]
23. Size of conclusion	.458
18. Allusions	.444
24. Paragraph linkage	−.439[b]
92. Conventional spelling	.436[b]
22. Conclusion	.413[a]

Factor 3 entails the logical expansion or elaboration of ideas. The variables of organization here (Measures 05, 07, 08) describe an ability to arrange logically more and more complex ideas as they expand dynamically. In impromptu writing, conclusions often grow in length along with logical elaboration of the body of the essay. Attention to specifics (Measure 13) may aid in elaboration of ideas and also may focus on unique events or allusions to well-known persons, places, and historical events. Expressing this logical dynamic negatively is static repetition of ideas. Therefore identical cohesive ties (Measure 29) are negatively associated—as are explicit linkings of one paragraph to the previous (Measure 24), a feature commonly associated with essays that are arranged by simple partition of the topic. The factor supports the findings of Durst (1984) and Haswell (1991) that high school and college students change toward more complex elaboration of ideas and that reliance on safe, static, and logical organizations holds that change back. It is important to see that Factor 3 describes the elaboration of ideas rather than their specification (see Factor 4) or their support (see Factor 11). It is also important to note that in the second principal-component extraction run, which installed the variable of holistic rating (Measure 01), that variable fell under this Factor 3 and under no other factor, loading positively at .513. Elaboration of ideas has been surmised as a primary factor in holistic ratings (Henning & Davison, 1987; Perkins & Brutte, 1990; Raforth & Rubin,

1984; Sullivan, 1997). The factor is strongly associated with postgraduate writing, with undergraduates showing only moderate change in that direction.

Factor 4: *Expansion of Given Information*

61. Relative clauses	.636[a]
31. Reference cohesive ties	.564[b]
74. Final free modification	.558[a, b]
67. Nominal modification	.497[a, b]
13. Level 1 generalizations	.464[b]
20. Introduction	.460[a, b]
75. Adverbs	–.447[a]
81. Infinitives	–.428
76. Initial adverbs	–.414
21. Size of introduction	.404
64. Base clauses	.402

Somewhat enigmatic, Factor 4 represents what might be called expansion of old or given information. Cohesive ties of reference often connect a new utterance back to an old topic locus. To expand on that given locus, convenient tools are relative clauses, final free modification, and logical restriction of nominals through modification (Measure 67). Level 1 or most specific generalizations expand the locus semantically. On the other hand, adverbs and infinitives are more likely to signal a turn to a new locus—hence their negative loadings here. Apparently, the factor is actively under change during the undergraduate years. The choice of tools to express it, however, continues changing after college. The competent working-world writers use fewer adverbial constructions and reference ties and do not continue the undergraduate increase in relative clauses, relying more on free modification, nominal modification, and specific sentence topics (Measure 13).

Factor 5: *Span of Explicit Logical Cohesion*

26. Logical indicators of cohesion	.578
20. Introduction	–.512[a, b]
39. Noun adjuncts	–.503[b]
27. Nonadditive logical indicators	.467[a, b]
09. Essay unity	–.446[a, b]
99. Conventional punctuation of main clauses	.434
07. Size of second-level logical pattern	–.429[b]
10. Qualifiers	.407

Positively, Factor 5 seems to represent the breadth of text that the writer prefers to organize logically through explicit methods of cohesion. A narrow or local span, expressed positively by the factor, includes a preference for explicit, logical transitions (such as *therefore* or *however*), a carefulness of punctuation of independent clauses (Measure 99), a nicety of qualification, a focus on the specific, and (in the second and third factoring runs) a preference for exemplification and a precision in expressing syntactic parallelism. Features conveying a broader span, expressed negatively by the factor, are functioning introductions, overall essay unification, and logical coherence covering passages usually larger than a paragraph (Measure 07). The change from undergraduate to postgraduate moves toward explicit connection of larger or more expansive structures of ideas (see Factor 3). The significant group differences in Measure 27, the one active variable correlating positively with Factor 5, actually make the same point, because the juniors and, even more, the postgraduates show a rate of explicit logical transitions decreasing from the rates of the freshmen and sophomores; both undergraduates and postgraduates gravitate away from heavy reliance on local cohesion (Hagen, 1971; Haswell, 1989).

Factor 6: *Establishment of Logical Boundaries*

03. Rhetorical mode	−.608
59. Nominal coordination	.509[a, b]
28. Cohesive ties	.436[b]
72. Initial free modification	.425
76. Initial adverbs	.416
71. Free modification	.414[b]
85. Syntactic parallelism	.405[a]

Positively, Factor 6 may represent the impulse toward the establishment of logical or classificatory boundaries. The variable of rhetorical mode, here with a negative loading, gave a higher rate to essays that qualified the topic as good or bad or that recommended a course of action, and a lower rate to essays that only defined or substantiated the topic (based on Eckhardt & Stewart, 1979). Logical classification is often the intent when nominals are coordinated, and initial nonrestrictive constructions usually serve to circumscribe place and time. Change in this factor is mixed. Although juniors show a decline in nominal coordination, the postgraduates record a substantial jump. Yet, the postgraduate rate of cohesive ties declines, and

their free modification rate increases. It is possible that the factor itself is relatively stable across these years.

Factor 7: *Nonrestrictive Expansion of Ideas*

71. Free modification	.564[b]
74. Final free modification	.503[a, b]
78. Final adverbs	.478[a]
84. Appositives	.421[a, b]
75. Adverbs	.419[a]
32. Substitution cohesive ties	.412
22. Conclusion	–.407[a]

With a nicely circumscribed set of variables, Factor 7 clearly describes nonrestrictive elaboration of ideas. It supports Christensen's (1968) argument that a distinct ingredient of professional written discourse is the topical expansion of new ideas outside the main clause, using semantically and syntactically free structures (Christensen & Christensen, 1967). The association with cohesive ties of substitution is not so curious when one thinks how often a logical progression by analogy or similarity is expressed by such ties (e.g., *so, likewise,* etc.). In further support of Christensen (1968), the evidence here strongly argues that nonrestrictive elaboration is under active increase during and after the college years.

Factor 8: *Fluency in Expression*

42. Passives	.571
05. Size of top logical pattern	.515[b]
21. Size of introduction	–.459
45. First-person singular	–.431[a, b]
98. Standard sentences	–.422
24. Paragraph linkage	.408[b]
25. Paragraph size	.401

Factor 8 represents fluency in the expression of ideas. Total essay length in words (Variable 02) was omitted from this first principal-components run, but when it was replaced in the third run, it appeared under this factor. It loaded positively, but only at .424. Evidently Factor 8 in the present run represents not only brute speed in getting words down on paper, but, more comprehensively, an ease in flow and linkage of ideas (Haswell, 1991, chap. 8). The passive usually aids flow of thought, and the sentence fragment (counted negatively in Variable 98) often reflects a colloquial fluidity of expression. Long and sometimes overambitious introductions often precede short,

blocked essays. Evidence for change in this component looks stronger if we add the fact that undergraduate groups record a consistent trend of greater essay length with age and the postgraduate writers a statistically significant jump, with essays that were more than 100 words longer than those of the entering freshmen.

Factor 9: *Breadth of Generalizations*

16. Level 4 generalizations	.650
14. Level 2 generalizations	−.570[a, b]
32. Substitution cohesive ties	−.498
29. Identical cohesive ties	.434
04. Top-level logical pattern	.405[a, b]

Factor 9 describes what might loosely be called *the power of inclusion*, an ability to collect and connect large bodies of information. Level 4 generalizations are subjects of t-units standing at the most abstract end of a four-tier classification of inclusiveness of category ("vegetation" is Level 4, "willow" Level 2; "physical appearance" is Level 4, "shirts" Level 2). With Factor 9 the contrast is repeated in the area of explicit cohesion, with identical ties tending to expand a given generalization and substitution ties tending to displace or move from one generalization to another. This factor shows little change.

Factor 10: *Facility With Vocabulary*

30. Synonym cohesive ties	.581
40. Elegant variation	.497
37. Infrequent words	.411[a, b]

Factor 10 suggests a facility or creativity with vocabulary. "Infrequent" words occur in print 10 times or less per million (Thorndike & Lorge, 1944), and the variable was one of the strongest indicators of undergraduate change toward professional performance, a change that Newberry (1967) was one of the first to document.

Factor 11: *Clarification*

11. Exemplification	.550[a]
96. Pronoun agreement	.422

This factor, and those remaining, require speculation more than interpretation, with few variables to work with and those few loading weakly. Factor 11 may touch on a rhetorical sense of clarity, which is the usual function of the explicit examples that were counted in

Measure 11, and may lie behind the attention to exact pronoun use. There is little change between groups.

Factor 12: *Whole Essay Organization*

05. Size of top logical pattern	.479[a, b]
82. Adverbial participles	−.439

The second extraction run added size of conclusions, with a positive loading of .426 (and with a statistically significant increase after the freshman year). This suggests the factor represents a kind of facility with whole essay logical organization, because the longer conclusions tended to fulfill the logical progression of the body of the essay.

Factor 13: *Balance of Ideas*

85. Syntactic parallelism	.415[a]

The second- and third-extraction runs added complexity of the second-level logical organization (Variable 6), loaded negatively, suggesting that this component may involve an organizational concentration on stasis or balance to the detriment of progressive thought. The significant change in syntactic parallelism is the junior group producing a lower rate than freshman and sophomore groups.

Factor 18: *Sophistication of Style*

95. Conventional predication	.404[b]

The second- and third-extraction runs added a positive association with medial adverbs and with appositives. Both are rather advanced writing maneuvers, suggesting that the factor may have to do with a stylistic grasp of sophisticated, conventional expression. All three variables show a marked increase with the postgraduate writers.

STEP 2: SELECTION OF MEASURES

This factoring does not claim that these particular 14 factors are the only components, or even the main components, of change in collegiate writing. Any factoring will be shaped by the particular cast of the variables selected for use. Nor does the factoring necessarily rank its variables according to their importance in understanding writing change during college. For instance, Haswell (1991) argues that

Table 2
Developmental Import of 14 Factors Extracted From a Principal-Components Factoring of Measures of Writing

	Statistically Significant Evidence of Progress	
Factor	1st to Junior Year	Undergraduate to Postgraduate
1. Span of Multiclause Units	Yes	Yes
2. Subclause Span of Bound Ideas	Yes	Yes
3. Logical Elaboration of Ideas	Yes in part	Yes
4. Expansion of Given Information	Yes	Yes
5. Span of Explicit Logical Cohesion	Yes	Yes
6. Establishment of Logical Boundaries	Partly	Yes in part
7. Nonrestrictive Expansion of Ideas	Yes	Yes
8. Fluency in Expression	Trend	Yes
9. Breadth of Generalizations	None	None
10. Facility With Vocabulary	Yes in part	Yes in part
11. Clarification	None	None
12. Whole-Essay Organization	Yes	Yes
13. Balance of Ideas	Yes	None
18. Sophistication of Style	None	Yes

prepositional use, variance of t-unit length, and complexity of top-level logical organization are significant accomplishments of undergraduates in terms of later postgraduate competence; yet in this factoring, these three measures are partialed away to subordinate roles. The 14 factors are just the main statistically noncorrelated components underlying this particular array of 84 variables.

But that fact carries it own weight. In the first place, study of the 14 factors with their associated variables allows a much more substantial answer to the question of whether writing change during the college years proceeds on a few discursive fronts or on many. The results of the analysis now show a broader, more multilateral change than was indicated even by the top-loading variables (compare Table 2 with Table 1). Factors range from bound and free modification of ideas through logical organization, coherence, specification, generalization, vocabulary, and stylistic facility. In the second place, the factoring can upgrade an individual variable as a measure of writing changes. It does so by allying it with other variables in the componential packet, variables that intuition and research more clearly associate with writing change. A facility in forming correct

parallel syntactic structures, by itself, does not seem of much importance as an indicator of writing growth. But Factor 3 (Logical Elaboration of Ideas) finds correctness of parallel structures associating positively with other competencies: the organization and expansion of points within topics, the inclusiveness of information within logical categories, and the reference to specific and concrete things and events. Together the variables constitute features of an important skill on the move during college, namely logical, grounded elaboration of ideas. In the third place, the factoring allows a choice from among variables representing a factor, according to goals of particular scholarly projects.

The goal of the present study is to document normal longitudinal changes in the writing of students during the first five semesters of college, changes associated with the performance of competent postgraduate workplace writers. Consequently, it follows four guidelines for selection of measures:

1. To take as much advantage from the factoring as possible, one measure will be taken from each of the factors.
2. But if a factor is too problematical, or if a feasible measure with a reasonable loading cannot be located, then the factor will be omitted.
3. In part, a measure will be feasible in terms of application, because in the end it will have to be applied to 128 essays—2 for each of the 64 participants.
4. A measure will best have a record of documented change in undergraduate writing toward postgraduate professional levels, ideally a record supported by a variety of research.

These guidelines helped choose one measure from each of nine factors. From Factor 1 (Span of Multiclause Units), mean length of sentence in words (Variable 53) was chosen instead of t-unit length, because sentences can be counted by computer and because in Haswell (1986) sentence length showed a statistically significant jump with juniors and with postgraduates, whereas t-unit length from freshman to sophomore recorded only a trend. Hunt (1965) preferred t-unit over sentence as a measure of "syntactic maturity," because he found fourth graders writing longer sentences than sixth graders, but he also recorded a steady increase in sentence length with sentence-combining exercises from eighth graders through to professional writers (Hunt, 1970). In impromptu essays, Newberry (1967) found an increase in length of sentences from grammar school students to advanced college students to postgraduates.

From Factor 2 (Subclause Span of Bound Ideas), mean length of clause in words (Variable 55) was chosen, as the highest-loading variable in the factor and as a measure with strongly documented evidence of activity from college to postcollege (e.g., Gebhard, 1978; Haswell, 1986; Hunt, 1970; Kerek et al., 1980; Stewart, 1978).

From Factor 3 (Logical Elaboration of Ideas), holistic score (Variable 01) was chosen. The highest loading variables in the first principal-component factoring—logical organization, syntactic parallelism, and cohesive ties—require rather sophisticated and lengthy analysis. When holistic scoring was included in the second factoring, it appeared in this factor alone (with a loading of .534). Although costly to apply, it is frequently used as a measure of writing performance. It also provides a synthesizing and semantic judgment, supplementing other measures that may be construed as counts only of discrete stylistic traits. As a measure of change in unrehearsed collegiate writing, holistic scoring has a mixed record. Using the system, Graham (1987) found no significant difference between freshman and senior, Scharton (1989) found a decline from freshmen through undergraduate tutors to graduate teaching assistants, and Freedman (1984) found no significant difference between college students and professionals. On the other hand, the system recorded a rising trend from freshman to junior and a significant increase with competent postgraduates in Haswell (1986), a significant rise in content and organization from freshmen to upperclassmen in Schumacher et al. (1984), and an increase in argument from freshman to senior in Whitla (1980).

From Factor 4 (Expansion of Given Information), percentage of words devoted to final free modification (Variable 74) was picked because it recorded significant increase with both undergraduates and postgraduates in Haswell (1986) and because it has a long record of implication in college and postcollege writing change (Christensen & Christensen, 1967; Faigley, 1979; Gebhard, 1978; Watson, 1983; Wolk, 1970).

From Factor 5 (Reliance on Local Cohesion), portion of total words devoted to the introduction (a new measure) was chosen as a proxy to represent the aspect of this factor that is active during the undergraduate years, although the aspect loaded negatively. Other measures loading highly in Factor 5, such as the logical transition markers (Variables 26 and 27) and the noun adjuncts, require time-consuming analysis. The measure of introduction length could be applied quickly with essays on disk. This choice of measure, however, takes a risky

shortcut, because the variable used in the factoring (Variable 20) was not word length of introduction but the sum of a primary-trait analysis of quality of introduction. Moreover, there is little past research on introductions and writing change. However, on quality of introduction (Variable 20), Haswell (1986) found a consistent trend upward from freshman to junior and a significant jump further to postgraduate writing.

From Factor 6 (Establishment of Logical Boundaries), nominal coordination was chosen, or more exactly, mean times noun structures are coordinated per clause (Variable 59). This choice also takes a risk, because the variable was one of the few in Haswell (1986) that recorded a regression during the undergraduate years: compared to freshmen, sophomores and juniors register a statistically significant decrease in use of nominal coordination, whereas the occupational writers record a strong increase. The other two high-loading variables in Factor 6, rhetorical mode and cohesive ties, are difficult to score. On the other hand, identification of coordinated noun structures is reliable, and clauses will already be counted for Factor 2. Clause rather than t-unit or sentence was chosen as denominator on the prediction that it would be the most stable of the three units among groups.

From Factor 7 (Nonrestrictive Expansion of Ideas), percentage of words devoted to all forms of free modification (Variable 71) was chosen as the highest loading variable, but also to avoid duplication with the measure of final free modification picked to represent Factor 4. In comparison with initial and medial, final free modification serves distinct composing and rhetorical purposes (Christensen & Christensen, 1967; Haswell, 1991, chap. 9; Williams, 1979). In the previous cross-sectional study (Haswell, 1986), sophomores and postgraduates record an increase in total free modification over 1st-year levels. Other research has found parallel changes in postgraduate writing (Christensen & Christensen, 1967; Freedman, 1984; Gebhard, 1978; Wolk, 1970).

From Factor 8 (Fluency in Expression), total essay length in words (Variable 02) was chosen, only in part because it is countable by computer. It appeared under Factor 8 when it was added for the second factoring run, loading at .424. In the cross-sectional study, the three undergraduate classes record a consistent trend upward in essay size, and the postgraduates a statistically significant jump further, a 33% increase in length over 1st-year output. Newberry (1967) found postgraduates producing more words in impromptu essays than did

secondary and undergraduate students. Graham (1987) found seniors writing longer essays than did freshmen, but not significantly so.

From Factor 10 (Facility With Vocabulary), the percentage of words in the essay which were nine or more letters long (Variable 38) was chosen to represent this factor of vocabulary facility. Because it could be calculated by computer, it served as a proxy for the painstaking count of infrequent words from Thorndike and Lorge (1944) (Measure 37). In Haswell (1986), the measure of words nine letters or more showed an increase (from freshman to sophomore and from undergraduate to postgraduate) statistically even more significant than that recorded by the Thorndike and Lorge measure. Findings from other studies record parallel changes, for instance in rate of "rare" words from Thorndike and Lorge from high school through college to postgraduates (Newberry, 1967), in formality of diction from high school seniors to college juniors (Freedman & Pringle, 1980), and in "diction" from college students to professional writers (Freedman, 1984).

Factors 9 (Breadth of Generalizations), 11 (Clarification), 12 (Whole Essay Organization), 13 (Balance of Ideas), and 18 (Sophistication of Style) were rejected as too problematical, too time consuming to apply, or as showing no undergraduate change in the cross-sectional study.

STEP 3: APPLICATION

As a first trial of the factoring, these nine measures were applied to a new and longitudinal sample of undergraduate writing.

Method. At Washington State University, since 1991 all entering students have been required to write a 2-hour impromptu essay for placement into 1st-year composition courses, and since 1993 all students with at least junior standing have been required to write the same kind of essay for placement into an upper-division composition course. By 1997, there were around 4,000 students with 1st-year and junior-year essays on file. This stock of writing was drawn on for the study.

The prompts are the same for the 1st-year and the 3rd-year testing. Students are given a paragraph-sized quotation and asked to respond to it along the line of inquiry suggested by one of four rhetorical frames (see Appendix B for one full prompt). Both quotations and

frames are rotated so that juniors rarely receive the same prompt they got 2 years previously. Writing conditions and administrative protocols are the same for the two exams. Rhetorical motive is roughly the same, because in both instances students are writing to achieve a desired placement or to avoid an undesired one. At least 2 years elapse between exams, so there should be minimal carryover effect.

All student essays are filed, and a sizable amount of information about the student at test-taking time is entered into a computer databank. This made selection of participants for the study, and collection of the writing, expeditious. A total of 64 students were chosen—a random sample large enough for group judgments and small enough for analysis. Sampling for the study began with the pool of students who had taken the 3rd-year examination, and randomly selected for students who had earned between 60 and 65 hours of credit, who had not transferred college credit, who were not second-language writers, and who had taken the 1st-year examination no more than 2½ years earlier.

Random selection was also used to control for the effect of rhetorical task on the writing. Of the four rhetorical frames used in the examinations, the two chosen were most equivalent in terms of placement outcomes, as determined by validation studies (Haswell, 1995). The two tasks, Frame 1 and Frame 3, were also free of significant gender differences.

> Frame 1: Read the following passage carefully. It conveys opinions with which many people may well disagree [paragraph quotation]. Clearly, there are other reasonable views on this complex issue. How do you, personally, resolve the differences among these opinions? Somewhere in your essay, we would like a description of the main issue, a summary of [the author's] view of it, and a comparison of that view with at least two other views.
>
> Frame 3: Read the following passage carefully. It may well give a misleading picture of the issue [paragraph quotation]. Clearly, it is easy to simplify this complex issue. Please discuss how the issue could be presented more accurately or honestly or fully. Somewhere in your essay, we would like a description of the main issue, a summary of [the author's] view of it, and a comparison of that view with at least two other views.

Random selection of participants continued until there was a sample of 32 students who had written on Frame 1 as entering students and

Frame 3 as juniors, and 32 students who had written on Frame 3 as entering students and Frame 1 as juniors.

Possible gender effects were controlled by selecting so that each of these two groups were composed half of females and half of males.

Essays were typed as written, using word-processing software capable of counting characters and words. Hard copies were printed for hand analysis, replacing the author's name with an identification code that hid 1st-year or junior-year status. Only one essay had to be replaced (exam booklet pages were missing). Otherwise, there was no attrition of the sample. It is important to realize that the population sampled in this longitudinal study was not entering students, but rather students who had lasted into their junior year. For the freshman classes covered by this study, 1991 to 1993, WSU retained 73% of entering students into their junior year.

Analysis. Three of the eight measures were applied by computer, and the rest by hand. All hand analysis was done by the author, except for the holistic rating, which was performed by a corps of writing instructors trained by the author.

Measure 1: Mean length of sentence in words, counted by computer. A sentence was taken as any word string ending in a period, interrogation mark, or exclamation point, as punctuated by the writer.

Measure 2: Mean length of clause in words. Definition of clause follows Hunt (1965, p. 15).

Measure 3: Holistic rating. Mean of four independent rates, on a scale ranging from 1 (*low*) to 8 (*high*), following a rubric giving equal weight to ideas, organization, support, sentences, vocabulary, and mechanics (Haswell, 1986). Raters judged typed essays unaware of the nature of the study and blind to name or academic class of writer.

Measure 4: Percentage of total words in the essay devoted to final free modification. Method of analysis followed Christensen (1968).

Measure 5: Percentage of total words in the essay devoted to the introduction. *Introduction* was taken to include the initial portion of an essay leading up to and including the writer's announcement of the "main issue," as identified by the prompt. Although intuitive, this understanding of introduction rarely proved problematic in application. Only 3 of the 64 essays did not signal the end of the introduction, so construed, with a paragraph break.

Measure 6: Mean number of coordinated noun structures per clause. This measure of nominal coordination followed Haswell (1986, p. 30). Verbal nominalizations were treated as noun structures. For example, "What this code is and why it exists will be discussed" was counted as an instance of nominal coordination.

Measure 7: Percentage of total words in the essay devoted to all kinds of free modification—initial, media, and free. Method of analysis followed Christensen (1968).

Measure 8: Length of essay in words, counted by computer. Hyphenated compounds, acronyms, and numbers were counted as single words.

Measure 9: Percentage of words nine or more letters long, counted by computer. Parts of hyphenated compounds were analyzed separately.

Findings. The alpha or pooled-rater reliability (Winer, 1962) of the four holistic scorers was .86.

Comparison by ANOVA of the two groups randomly selected for the two different orders of rhetorical task supported a working equivalence of the groups. There were no significant differences between the group who wrote Frame 1 first and Frame 3 second ($n = 32$) and the group who wrote Frame 3 first and Frame 1 second ($n = 32$), in terms of high school grade point average (GPA), accumulated college GPA as juniors, admissions index, or amount of wait time between the 1st-year and the 3rd-year exams. Analysis of these two groups by ANOVA also found no significant differences in outcomes with any of the nine measures. That is, the order of rhetorical prompt was not associated with the changes students recorded on these nine measures of writing.

Across these groups, that is, across all 64 students, significant longitudinal change was recorded in eight of the nine measures of writing. Confidence testing of the difference between 1st and 3rd year performance was accomplished by Spearman correlated t tests. Correlated pairs were each student's performance on the matriculation exam and on the junior exam. Table 3 displays the results. All measures but the coordination of nominals achieved an acceptable level of confidence ($p > .05$). Also, as Table 4 shows, all nine measures show a change from 1st to 3rd year in the direction that professional, postgraduate writing recorded in comparison with undergraduate writing in Haswell (1986).

DISCUSSION

With what confidence can one generalize the main finding in this study—that a selection of juniors have changed in their writing since freshmen toward postgraduate levels in eight of nine fairly distinct measures of writing? The random sample of 64 participants seems to

Table 3

Longitudinal Data on Nine Measures of Writing: Comparing the Essays of 64 Students Composed at Their 1st and Junior Years

	Spearman Correlated t Test			Freshmen		Juniors		
Measure	*t*-Value	Confidence (*p*)	Correlation	*M*	*SD*	*M*	*SD*	Change Toward Postgraduate Performance
1. Mean length of sentences in words	3.06	.003	.483	17.50	4.01	18.91	3.08	+1.41 words
2. Mean length of clauses in words	2.75	.008	.399	9.02	1.73	9.66	1.61	+0.64 clause
3. Holistic rating mean (8-point scale)	4.96	.000	.406	3.71	1.50	4.72	1.55	+1.01 scale points
4. Words in final free modifiers (%)	5.24	.000	.011	7.4	4.4	11.6	5.0	+4.2 total words
5. Words in introduction (%)	2.25	.028	.128	14.5	7.8	17.4	8.4	+2.9 total words
6. Mean coordinated nominals per clause	1.04	.302	.415	0.14	0.06	0.15	0.09	+0.01 per clause
7. Words in free modifiers (%)	5.87	.000	.466	19.6	7.0	24.9	7.3	+5.3 total words
8. Length of essay in words	7.21	.000	.428	400.0	127.2	548.9	169.7	+148.9 words
9. Words ≥ nine letters (%)	2.23	.029	.030	7.5	2.8	8.3	2.5	+0.8 words

Table 4
Evidence of Improvement on Nine Measures of College Writing From Four Groups in Two Studies

	Study 1				Study 2		
	Group 1		Group 2		Group 3		Group 4
Measure	FY1	JY3	FY3	JY1	FY	JY	PG
1. Length of sentence in words							
M	16.92	17.81	18.08	19.99	17.2	18.6	19.8
SD	3.8	2.9	4.2	2.8	3.3	4.3	3.6
2. Holistic score							
M	3.63	4.41	3.79	5.04	3.86	4.24	5.70
SD	1.2	1.6	1.8	1.4	1.3	1.5	1.7
3. Words in final free modification							
Percentage	6.7	10.8	8.0	12.3	5.5	8.2	10.8
SD	4.4	4.7	4.4	7.0	5.7	5.9	6.2
4. Amount of essay devoted to introduction							
Percentage	14.1	17.4	14.9	17.4	6.5[a]	6.3[a]	6.7[a]
SD	7.6	8.0	8.1	8.9	NA[b]	NA[b]	NA[b]
5. Clauses with coordinated nominals							
Percentage	12.8	14.1	14.7	15.8	11.0	6.1	18.4
SD	5.2	8.8	7.8	9.0	8.4	7.0	10.4
6. Words in free modification							
Percentage	17.8	22.5	21.4	27.3	19.2	19.0	26.2
SD	7.4	7.0	6.2	7.0	7.6	7.9	8.0
7. Length of essay in words							
M	399.1	515.9	401.0	581.9	316.9	364.4	433.3
SD	129.9	151.2	126.6	182.9	118.9	131.2	152.3
8. Words ≥ nine letters							
Percentage	7.7	8.6	7.2	8.1	6.2[c]	7.5[c]	9.9
SD	2.4	3.1	3.2	1.7	2.1	2.7	3.5
9. Length of clause in words							
M	8.74	9.68	9.31	9.64	9.1	9.1	10.2
SD	1.4	1.6	2.0	1.7	1.8	1.3	1.5

NOTE: For each group, $n = 32$. Study 1 involves students writing on Rhetorical Task 1 as freshmen (FY1) and Rhetorical Task 3 as juniors (JY3); Group 2 consists of students writing on Rhetorical Task 3 as freshmen (FY3) and Rhetorical Task 1 as juniors (JY1). Group 3 consists of students, some 1st year (FY) and some junior year (JY), writing on the same rhetorical task at the same point in time; Group 4 consists of employed post-graduates (PG) writing on the same task as Group 3 (Haswell, 1986).
a. The measure of essays' introduction was somewhat different in Study 1 and Study 2.
b. Variance cannot be calculated because the percentage is a derivation from previous data.

compare well with the whole population of juniors at WSU. The sample's cumulative GPA was 3.23, compared to 3.19 for their junior cohorts as a whole. Their distribution by academic major matches well the distribution for all juniors standing for the junior-level placement examination. So does their record of achievement on the exam: 7.8% of the sample and 10.2% of all juniors were required to take more coursework; 81.9% of the sample and 79.0% of all juniors earned a simple pass; 10.3% of the sample and 10.8% of all juniors earned distinction.

How well this sample represents midcareer college students in general is, of course, a different question, and one this study cannot answer. It also does not control for cohort effects. If cross-sectional sampling opens the door to uncontrolled cohort effects because it selects different historical age groups at the same point in time, longitudinal sampling restricts its gaze to one cohort and thereby ducks the issue. The 64 students sampled here actually come from 3 matriculation years (1991-1993), a fact that should help dissipate cohort effects of any one of those years, but together the years may reflect one generation of students whose performance on the two tasks another generation may disqualify. In study of human change, cohort effects cannot be fully controlled even by combining cross-sectional and longitudinal samplings (Adam, 1979; Aiken, 1998). The same is true for task effects. Counterbalancing two rhetorical prompts may control for the two in terms of each other, but it says nothing about other possible prompts. Both of the two tasks used may tap into special experiences of the older students having little to do with writing skill per se.

Group norms also hide individual differences. If eight of the nine measures show statistically significant change for the group, still few of the students in the sample show such wholesale advance across all eight of the measures, and some record hardly any change from freshman to junior year. Of the 64, 15 received a lower holistic rating on their junior essay. Most of these 15 students had written better than average essays in their 1st year. The mean holistic rate on their original essays was 5.08, compared to a mean of 3.29 for the other 49 students and a mean of 3.71 for all 64. Statistical regression may account for some of this difference, but it is also possible that students who write comparatively well on entering college do not find as much of an incentive to change in their writing in their first 2 years of courses.

Clark (1987) found a similar phenomenon in her case study of six juniors with 1st and 3rd year impromptu essays.

In short, the extent to which the group findings can be called *normative*, that is, expected to hold under different conditions and contexts, can only be discovered by running comparable studies under different conditions and contexts. But some answer is available already, much of it in the positive. Previously cited are past investigations under a wide variety of situations finding comparable changes in essay length, sentence length, clause length, free modification, learned vocabulary, and holistic score (though evidence for the last is mixed). Comparison with just one study (Haswell, 1986) provides especially telling support because so many of the measures were replicated. Table 4 shows that on the nine measures, the shift from 1st to 3rd year, and the direction of that shift toward competent postgraduate writing, is very consistent, cutting across six different matriculation classes, four different rhetorical prompts, and two different methods of group comparison (longitudinal and cross-sectional). The only exceptions are the junior performances in the earlier study on the measures of introduction, coordinated nominals, and combined free modification.

It is worth noting that in the current study, analysis by ANOVA found none of these nine measures of change from 1st-year to 3rd-year performance associated with the gender of the writer. Table 5 breaks down the findings by gender and shows that with every measure, both genders recapitulate the group changes. The leveling of gender differences in performance suggests another way that these longitudinal writing changes can be construed as normative.

It is to raise another issue entirely, however, to ask how well the cluster of measures worked as measures of writing. As an index of undergraduate writing change, the measure of compounded nominals (Measure 06) looks questionable, because it did not support the earlier finding of junior change (Haswell, 1986). On the other hand, evidence of change was recorded by the two proxy measures, the word length of introductions (Measure 05), and the letter length of words (Measure 08). A second question is how well these, or any of the nine measures, functioned as representatives of their respective factors. It is a question that cannot be answered satisfactorily at this point. Representing a multivariable factor with one variable always entails loss of descriptive power. The most doubtful of measures in this sense is the holistic rating (Measure 03). It represented Factor 3 (Logical Elaboration of Ideas), which would have been better served

Table 5
Differences Between Male and Female Writing Performance: Comparing the Essays of 64 Students Composed at Their 1st and Junior Years

	Males ($n = 32$)				Females ($n = 32$)			
	Freshman		Junior		Freshman		Junior	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1. Mean length of sentences in words	17.68	4.04	18.69	3.47	17.31	4.04	19.14	2.66
2. Mean length of clauses in words	9.32	2.05	9.62	1.67	8.73	1.30	9.70	1.58
3. Holistic rating mean (8-point scale)	3.46	1.58	4.81	1.57	3.95	1.39	4.63	1.54
4. Words in final free modifiers (%)	7.8	4.6	12.1	4.6	6.9	4.2	11.0	5.3
5. Words in introduction (%)	13.5	7.5	16.7	8.3	15.5	8.1	18.2	8.6
6. Mean coordinated nominals per clause	0.14	0.07	0.14	0.08	0.14	0.06	0.16	0.10
7. Words in free modifiers (%)	20.0	7.4	25.6	7.1	19.2	6.7	24.2	7.6
8. Length of essay in words	380.1	128.7	559.7	166.6	419.9	124.6	538.2	174.8
9. Words ≥ nine letters (%)	7.6	3.1	8.5	2.5	7.3	2.4	8.2	2.5

NOTE: All differences are not significant by Spearman correlated *t* test.

by a more direct measure of logical organization. The holistic, however, did add its unique contribution to the set of factors. A stepwise multiple regression run on the holistic rating as dependent variable found the other eight measures predicting only 57.9% of the rating's contribution to total variance.

The degree of redundancy of measures raises a third question, one more amenable to analysis: How much does this selection from the original factoring, this new set of measures, avoid multicollinearity? Correlation matrices of the nine measures (see Table 6) offer an answer. The large majority of correlations between variables are quite weak. Of the 36 possible correlations, only 4 consistently correlate at .5 or higher: sentence length (Measure 01) with clause length (Measure 02), clause length (Measure 02) with coordination of nominals (Measure 06), holistic rating (Measure 03) with word length of essay (Measure 08), and final free modification (Measure 04) with combined free

Table 6
*Pearson Product-Moment Correlation Matrices on Nine Measures Applied to Two Sets of Student Essays (*N = *64)*

Measure	1	2	3	4	5	6	7	8	9
1. Mean length of sentences in words									
1st year	—								
3rd year	—								
2. Mean length of clauses in words									
1st year	0.605	—							
3rd year	0.368	—							
3. Holistic rating mean (8-point scale)									
1st year	0.087	0.019	—						
3rd year	0.258	0.336	—						
4. Words in final free modifiers (%)									
1st year	0.217	0.192	0.066	—					
3rd year	0.205	0.252	0.250	—					
5. Words in introduction (%)									
1st year	–0.137	–0.208	–0.020	–0.042	—				
3rd year	0.251	0.019	–0.194	–0.068	—				
6. Mean coordinated nominals per clause									
1st year	0.233	0.552	–0.001	0.309	–0.085	—			
3rd year	0.067	0.437	0.039	–0.013	0.135	—			
7. Words in free modifiers (%)									
1st year	0.262	0.225	0.167	0.508	–0.093	0.148	—		
3rd year	0.377	0.115	0.281	0.564	–0.005	–0.078	—		
8. Length of essay in words									
1st year	0.101	–0.091	0.717	0.106	0.103	–0.238	0.090	—	
3rd year	0.055	0.131	0.520	0.197	–0.391	–0.006	0.044	—	
9. Words ≥ nine letters									
1st year	0.232	0.426	0.028	0.127	–0.149	0.399	0.164	–0.050	—
3rd year	–0.045	0.404	0.356	0.105	–0.039	0.123	–0.083	0.055	—

modification (Measure 07). Such a low degree of redundancy endows this cluster of measures with added authority as an indicator of change in college writing.

DIRECTIONS FOR FURTHER RESEARCH

Clearly, for researchers exploring normative change in the writing of college students, multiple factors is a key to future advance. To index such change, studies of text along the lines of the present one may locate more measures, and better sets of measures, but until variables include a broader range of venues and concerns, the work will lie quiescent. Although the present study found a rise in holistic score longitudinally from freshman to junior, and movement along a number of rhetorical fronts toward competent occupational writing, those findings in themselves cannot provide definitive answers to crucial questions. Are these changes part of a broad writing competence, applicable to different writing situations and purposes? Do they truly constitute improvement in writing? Are they part of normal adult maturation or development of expertise? Can they be attributed to a particular curriculum? What part do they play in the social, cultural, and internal lives of the writers? What necessary part, if any, do they play in the life of any one of the 64 writers? To bring the current study to bear on these and similar kinds of questions, and, more broadly, to help move Kitzhaber's (1963) project out of its current state of near dormancy, will mean expanding the investigation into further research terrain. Six territories show particular promise, though by no means do they exhaust the possibilities.

Contextual. The concept of normative requires that some effect hold generally across not only differences in individuals but also differences in context. Is it normal for college students, as they continue in college, to write in larger and larger packets of syntactically bound ideas? To answer, we need to see how that effect appears in a variety of contexts—different rhetorical tasks and genres, composing processes and pressures, target audiences. It has been argued that because stylistic features vary with rhetorical context, research studies cannot use them unproblematically as indicators of writing "growth" (e.g., Faigley, 1980; North, 1987). But it can also be argued that how writers vary their writing in response to different contexts is part of writing change. If an effect consistently appears in some contexts and not in

others, that does not invalidate the hypothesis of normality but shapes it into something more complicated and accurate. Much basic information about change in student writing can be located simply by varying the situations in which it is elicited. It is arguable that research will advance more by exploring new contexts with old variables and old contexts with new variables than it will with new variables in new contexts.

Valuative. However, even in the unlikely case that a formal language change across time operates consistently in all contexts, questions of value still remain. Language change in human lives is not indexed directly by language production but mediated through social value systems, as are other human givens such as class, race, and gender (e.g., Ochs, 1993). One of the most consistent, cross-context changes in college writing, according to the literature, is the lengthening of sentences. But longer sentences may be good or bad. The motive to increase sentence length may be creditable (e.g., to convey the complexity of a topic) or reprehensible (e.g., to show off), and reader response to longer sentences may be positive (e.g., sounds mature) or negative (e.g., is confusing). This problem in indexicality—the way value complicates any direct association of language feature and change across time—lies behind a common critique of the classic multivariate studies of writing change by Hunt (1965, 1970), Loban (1976), and Britton, Burgess, Martin, McLeod, and Rosen (1975), that their frequency-count measures of growth are fundamentally equivocal (e.g., Bereiter, 1980; Faigley, 1980).

What will dissipate the equivocality are studies that connect language performance with value systems and evaluation systems. So the current factoring not only associates sentence length with certain other multiclause units, such as greater variance in length of t-units, larger share of nominal modification placed after the noun, increasing reliance on prepositional strings and free modification, and avoidance of the generic *you* (see Factor 1); it also associates all these features with the performance of postgraduates who have been prejudged as competent writers. This longer "competent" sentence is not just any kind of longer sentence. It is a sentence serving specific rhetorical motives, opting for syntactic and tonal choices that heighten the register, generate rhetorical emphasis, and increase readability of thought units of a certain logical complexity. Further directions for study abound. Reader response experiments can determine if this competent sentence differs from the heavily nominalized sentence that Hake and Williams (1981) found English teachers liked, or from

the heavily adjectivized, passivized, and embedded sentence for which Raforth and Combs (1983) found 1st-year college students had no ear. Cognitive processing studies can discover how close it is to the active, subject-simple and object-complex sentence recommended by Williams (1981). Genre studies can ally it with the nonrestrictive, cumulative sentence championed by Christensen (1968). And in all of this, one of the most productive ways to establish value systems will be through student/expert comparisons of performance (a way the present study does not follow directly). The tasks of the two groups should be made as equivalent as possible, as in the studies of Faigley and Witte (1981), Freedman (1984), and Kaufer, Hayes, and Flower (1986).

Theoretical. In an influential essay, Faigley (1980) argues that the feature of syntactic length alone cannot define rhetorical maturity because some competent prose is written at grade-school norms (e.g., cooking recipes at fourth-grade levels of t-unit length) and some incompetent prose falls at the other end of the spectrum of length (e.g., bureaucratese). He concludes that maturity, and similar concepts such as growth, must be represented by measurements with more rhetorical import (e.g., measurements of propositions, coherence, pragmatics, and audience reception). Even such a description of textual accomplishment, however, will not satisfy terms such as *maturity* and *growth* fully. In the current study, even the holistic rating does not provide clear evidence of maturation from freshman to junior, although it surely appraised more than just syntax, because the holistic itself, as an evaluative system, needs to be questioned on the grounds of maturity. Freedman (1984) found writing teachers holistically rating professional essays no better than student essays, in part because the teachers were unable to appreciate the passion, feeling, and definiteness of the professionals. So were the teacher raters or the professional writers rhetorically more mature?

Any concept of maturity, growth, or improvement requires a base of theory to set the values that rescue measurement of change from ambiguity. Not only maturation (Cirillo & Wapner, 1986; Keil, 1987) but rhetorical effectiveness itself (Gee, 1994) are theoretical constructions, providing values, beliefs, claims to knowledge, and reasons for choice. Therein lie not barriers but avenues for research. Many theories offer explanatory paradigms for rhetorical growth, each different and each recommending different rationales, hypotheses, and designs for research. Consider again the longer sentences of the juniors. How can research test whether they represent advance or

regression? Theories of situated rhetoric (e.g., Selzer & Crowley, in press) would define discursive maturity as successful adaptation to concrete rhetorical settings, and suggest a testing of the freshman and junior sentences, embedded in their essays, against the judgment of academics who read the respective examinations. Theories of practitioner expertise (e.g., Schön, 1983) would define maturity as a reflective ability to adjust successfully when previous knowing hits a snag, and suggest think-aloud process experiments (e.g., Kaufer et al., 1986). Theories of postadolescent personal development have defined maturity in many ways. They might recommend a comparative scrutiny of freshman and junior sentences for dialectical thinking (Basseches, 1987) or for tendencies to jump and chain logical classificatory boundaries (Haswell, 1991, chap. 9; Richards & Commons, 1984). Sociolinguistic theories of language use (e.g., Bolinger, 1980; Witte & Flach, 1994) often define discursive maturity in part as avoidance of deception or self-deception, and might well hypothesize the lengthening of sentences as regression, recommending research designs using close semantic analysis or reader response. In sum, any research that wishes to move beyond a descriptive report of change toward notions of maturity, expertise, improvement, or development needs to bare its theoretical grounding and its application to design and findings. For the present study, that would require exploring, theoretically, the assumption that undergraduates should change rather than not change in their writing performance (an assumption that controlled the choice of measurement variables), and the assumption that workplace writing habits constitute a positive end for that change.

Curricular. Researchers will not be satisfied if a university administration—it is not unlikely—appropriates findings such as those of the present study and advertises them as evidence that the curriculum is successfully improving the writings of its students. Instead, researchers must take seriously White's (1989) position that a causal relationship between an instructional program and writing improvement is very difficult to establish. Historically, White has been supported by three common institutional experiences. Variables chosen for value-added studies reveal little or no evidence of improvement in writing across the undergraduate years (e.g., Curry & Hager, 1987; Graham, 1987; Scharton, 1989). Even when evidence is found, it cannot be attributed to any particular curriculum convincingly (e.g., Hurtgen, 1997; Vandament, 1987). And the handiest contrast groups, native

and transfer students, do not show statistical differences (Curry & Hager, 1987; Haswell, 1995).

Nevertheless, if the lessons in this history are respected, the relationship between curriculum and writing change still offers fertile ground for research. Impact of instruction on writing can be meaningfully sought, so long as it is tested close to the termination of the course, and so long as the variables chosen are tightly implicated in the instruction of the course (Banta, 1997). As for impact extending much beyond the course, common sense argues that influence of particular curricular programs will be inextricably dissolved into the entire educational experience, and that studies should be multifactorial with a substantial portion devoted to student experiences, opinion, self-rating, and self-reflection, as Witte and Faigley (1983) recommend, and as Sternglass (1997) at City College of CUNY, Light (1993) and his associates at Harvard, and Astin (1997) in his multiuniversity study demonstrate can bear good fruit. Appropriately, the findings of the present study form only one small part of an ongoing multiple inquiry into the composition and assessment program at WSU (Haswell, 1998). Interinstitutional comparisons are also a promising and untapped direction for research, so long as measurements again are multiple, and more interpretable than standardized tests with holistic ratings or item arrays that cannot be broken down (Astin, 1987). Evidence of instruction first and writing improvement second, of course, can never prove that the first necessarily had anything to do with the second. But it can shift the burden of proof. It can prove that instruction did not lead to regression or no gain, and therefore hands over to someone else the task to show that something other than instruction produced the betterment.

Methodological. In the study of undergraduate writing change, rewards lie not only with multifactorial approaches but also with multimethodological ones. The benefits of combining research methods have been much appreciated by social scientists (e.g., Brewer & Hunter, 1989) but only occasionally by compositionists (e.g., Charney, 1996). The benefits can be illustrated easily by noting some connections between ethnographic longitudinal inquiry into change during college and the present textual study. Haas (1994), Chiseri-Strater (1991), and Sternglass (1997), for instance, all find a major influence of academic domain on the evolving composing decisions of students, especially when they have decided on a major—an influence that can be allied with the present study's finding of an increase in learned vocabulary and in syntactic forms common to disciplinary prose such as

larger clauses and sentences, more elaborated nominals, and more free modification. Haas (1994) and Sternglass (1997) also record a growing metadiscursive awareness of the situatedness of authors and the diversity of readers, which may account in part for the increase in space given over to introductions and final free modification. Chiseri-Strater (1991), Haas (1994), Sternglass (1997), and Spack (1997), observing an English as a second language (ESL) writer, all discern a developmental leap when students find that college-level writing assignments allow them to promote their own thoughts and opinions instead of merely repeating those of authorities—a rhetorical motive that often finds a syntactic expression in free modification (cf. Haswell, 1991, pp. 239-240).

At the moment, these connections between quantified measures of written products and ethnographic insights into the psyche and situation of writers are speculation, but they are also reasonable hypotheses, ones that could be tested by a research design combining ethnography and text analysis. Sternglass (1997, p. 36) cites a long list of developmental writing traits from Haswell (1991), most of them inferred from quantified, cross-sectional group measures, and notes that she found qualitative, longitudinal, case study evidence of many of them in the writing, words, and behavior of her nine participants. Two research methodologies, often considered enemies, may actually complement each other to the degree that they exclude each other.

Individualistic. One productive area of complementarity is the intersection of group and individual. These case studies, with their access to the opinions, feelings, and daily lives of students, show that rhetorical progress during college can ebb, fragment, or drag for years, that writing change is always complex, conflicted, and contextual; in other words, it is always highly individual. The ethnographic insight is a valuable corrective to quantified study of groups, in which normative change so easily covers up eccentric individual change. By the same token, though, the construction of norms provides measures—often lacking or only to be inferred in ethnographic inquiries—by which the individual can be understood as singular. And although study of group norms may hide individual singularities, to the degree that it is quantified and to the degree that the investigation is multifactorial, it can also uncover them. For two instances from the present study, the nine-measure data display easily locates a student who doubled his holistic rating as a junior and increased in all the other measures except for length of sentences, which decreased an astonishing 6.7 words per sentence; and the student who increased on

all measures, including a jump of 242 words in the length of essay, with the exception of her holistic rating, which dropped from 5.5 to 2.0 on a scale of 8. The numbers beg for further study, both textual and ethnographic. In the area of individualistic change in writing competence during college nearly everything is waiting to be discovered, though support for the unique development of individuals is currently available from researchers in other disciplines, such as Johnstone (1996) in sociolinguistics and McAdams (1993) in life story.

This rough and ready survey of six directions for research—and there are plenty more routes—does more than just point out ways the present study needs to be continued before it can be of good use. It shadows forth an argument about the most productive means by which Kitzhaber's (1963) project of describing and understanding writing change during college can be carried out by a community of researchers. That means will not be a search for the one holy measurement, nor the definitive investigation, nor the ideal methodology. It will entail a lengthy interplay of variables, experiments, replications, and reviews conducted in an ever-growing and incremental set of contexts and populations. The payoff, however long it takes, will be a picture of customary undergraduate change in writing performance, descriptive yet connected to educational and occupational values, paradigmatic yet situated enough for researchers and teachers to set feasible hypotheses and expectations, specific yet flexible enough that students can find room for their unique personalities, potentials, and goals.

APPENDIX A
Variables Used in Three Principal-Component Factorings

Variable Number	Description
1[a]	Holistic score: 8-point scale
2[a]	Essay size: total words
3	Rhetorical mode: scale composed of definition, substantiation, evaluation, and recommendation (see Eckhardt & Stewart, 1979)
4	Top-level logical pattern: scale based on complexity of the logical pattern that organizes the largest part of the essay (see Haswell, 1986)
5	Size of top-level logical pattern: percentage of total words
6	Second-level logical pattern: same scale as in Measure 4 but applied to largest embedded part of essay
7	Size of second-level logical pattern: percentage of total words
8	Logical depth: largest number of logical embeddings

(continued)

APPENDIX A Continued

Variable Number	Description
9	Essay unity: where the top-level logical pattern embraces all the essay, percentage of essays in group
10	Qualifiers: expressions of qualification (*could, on the whole*), per word
11	Exemplification: instances of logical examples or illustrations, per word
13	Level 1 generalizations: most specific (my classmates), per t-unit
14	Level 2 generalizations: basic level (teenagers), per t-unit
15	Level 3 generalizations: most general (people), per t-unit
16	Level 4 generalizations: abstract (society), per t-unit
17	Unique events: at least one reference to a singular event, percentage of essays in group
18	Allusions: references to person, place, or historical event, percentage of essays in group
20	Introduction: 11-point primary-trait scale
21	Size of introduction: percentage of total words devoted to the introduction
22	Conclusion: 5-point primary-trait scale
23	Size of conclusion: percentage of total words devoted to conclusion
24	Paragraph linkage: paragraphs that connect explicitly with the previous paragraph, percentage of inside paragraphs in the essay
25	Paragraph size: mean words per paragraph
26[b]	Logical indicators of cohesion: explicit logical connectors (*thus, such as*), including additives (*also, moreover*), per word
27	Nonadditive logical indicators: instances of illustratives (*for example*), illatives (*hence*), adversatives (*yet*), and causatives (*since*), per word
28[b]	Cohesive ties: instances of all cohesive ties (Halliday & Hasan, 1976), per word
29	Identical cohesive ties: per word
30	Synonym cohesive ties: per word
31	Reference cohesive ties: per word
32	Substitution cohesive ties: per word
37	Infrequent words: words occurring in print 10 times or less per million (*baggy, discomfort, nuisance*), percentage of total words
38	Long words: words nine or more letters long, percentage of total words
39	Noun adjuncts: adjuncts (*dog collar, fashion show*), percentage of all prenominal adjectives
40	Elegant variation: portion of identical and synonymous ties that are synonymous ties (see Variables 29 and 30)
42	Passives: predicates that are passivized (*overcome, we are told*), percentage of total predicates
43	Process verbs: predicates that are nonstative (*fight*) rather than stative (*own*), percentage of total predicates
44	Pronominalization: pronouns, percentage of total references to humans or human groups
45	First-person singular: essays with more than a fourth of pronouns expressed as the first-person singular, percentage of essays in group

(continued)

APPENDIX A Continued

Variable Number	Description
46	Second-person plural: essays with more than a fourth of pronouns expressed as the second-person generic *you*, percentage of essays in group
53	Sentence size: mean length of sentences, in words
54	t-unit size: mean length of t-units, in words
55	Clause size: mean length of clauses, in words
56[b]	Clause/t-unit ratio: number of clauses divided by number of t-units
57	Coordination of t-units: percentage of sentences compounded of two or more t-units
58	Predicate coordination: number of times predicates were coordinated, per clause
59	Nominal coordination: number of times noun structures were coordinated, per clause
60	Subordinate clauses: percentage of total words occurring in subordinate clauses
61	Relative clauses: percentage of total words occurring in relative clauses
62	WH clauses: percentage of total words occurring in nominalizations formed from questions or exclamations (*How they acted was . . .*)
63	That clauses: percentage of total words occurring in nominalizations formed from statements (*That they always were acting was . . .*)
64	Base clauses: percentage of total words occurring in main clauses of multiclause t-units
65	Monoclauses: percentage of total words occurring in single-clause t-units
66	Independent clauses: percentage of total words occurring in main clauses
67	Nominal modification: percentage of total words occurring in structures modifying nominal heads
68	Variance in nominal complexity: number of different-sized strings of nominal modification (Hunt, 1965, pp. 114-115)
69	Postnominal modification: percentage of total words occurring in modifying structures after the noun head
70[b]	Pre/postnominal modification ratio: Amount of prenominal modification divided by amount of postnominal modification
71[b]	Free modification: percentage of total words occurring in nonrestrictive structures (Christensen, 1968)
72	Initial free modification: percentage of total words occurring in initial free modification
73	Medial free modification: percentage of total words occurring in medial free modification
74	Final free modification: percentage of total words occurring in final free modification
75[b]	Adverbs: percentage of total words occurring in structure modifying predicates or whole clauses
76	Initial adverbs: percentage of words in adverbial structures placed before the grammatical subject of the t-unit

(continued)

APPENDIX A Continued

Variable Number	Description
77	Medial adverbs: percentage of words in adverbial structures placed between the subject and the end of the t-unit
78	Final adverbs: percentage of words in adverbial structures appended to the end of t-units
79	Prepositions: percentage of total words in prepositional phrases
80	Prepositional strings: percentage of total words in prepositions composed of strings of two or more prepositional phrases
81	Infinitives: percentage of total words in infinitival structures, including crypto-infinitives (*help, make*)
82	Adverbial participles: percentage of essays in group using participial constructions modifying clauses (*In considering the question of beauty, we must . . .*)
83	Nominalized participles: percentage of essays in group using participial constructions serving as complement (*seen smoking the correct cigarettes*) or gerund (*a method for learning what is right*)
84	Appositives: percentage of essays in group using one or more appositives
85	Syntactic parallelism: occurrence of syntactic parallelism within clauses (Hiatt, 1977)
86	Dependent sentence openers: percentage of all types of sentence openers that are dependent on the grammatical subject
87	t-unit variance: standard deviation of word length of t-units
92	Conventional spelling: percentage of total words spelled according to *The American Heritage Dictionary*
93	Conventional possessives: percentage of essays in group with no errors in the formation of possessives
94	Conventional contractions: percentage of essays in group with no errors in the formation of contractions
95	Conventional predication: percentage of main verbs that express reasonable semantic agreement with the subject or complement
96	Pronoun agreement: percentage of essays in group with no errors in agreement between a human pronoun and its antecedent
97	Conventional parallelism: percentage of syntactic parallelism (see Variable 85) with coordinated elements of the same grammatical class
98	Standard sentences: percentage of essays in group with no sentence fragment
99	Conventional punctuation of main clauses: percentage of essays in group separating independent clauses with punctuation approved by standard stylebooks

NOTE: Variable numbers refer to an original set of 107 measures (Haswell, 1986). Eliminated here are measures that were not preset before analysis and measures that do not adjust for length of essay.
a. Variables eliminated from the first and third principal-components factoring.
b. Variables eliminated from only the third principal-components factoring.

APPENDIX B
Freshman and Junior Prompt for a 2-Hour Essay, Using Rhetorical Frame 1

YOUR PURPOSE

To show us your ability to write an essay at the level expected of students entering Washington State University [or of juniors at Washington State University]. Especially we are looking for an essay that is

- focused around a main point,
- organized throughout,
- supported persuasively, and
- proofread.

YOUR WRITING ASSIGNMENT

Read the following passage carefully. It conveys opinions with which many people may well disagree.

> I was talking with an Alaskan hunting guide who had killed some thirty-odd wolves himself from a plane, alone, and flown hunters who had killed almost four hundred more. As he described with his hands the movement of the plane, the tack of its approach, his body began to lean into the movement and he shook his head as if to say no words could tell it. For him the thing was not the killing; it was that moment when the blast of the shotgun hit the wolf and flattened him—because the wolf's legs never stopped driving. In that same instant the animal was fighting to go on, to stay on its feet, to shake off the impact of the buckshot. The man spoke with awed respect of the animal's will to live, its bone and muscle shattered, blood streaking the snow, but refusing to fall. "When the legs stop, you know he's dead. He doesn't quit until there's nothing left." He spoke as though he himself would never be a quitter in life because he had seen this thing. Four hundred times.
>
> Here is what I should have said to this guide: It does not demean men to want to be what they imagine the wolf to be, but it demeans them to kill the animal for it.
>
> —Barry Lopez, "Wolfing for Sport"

Clearly, there are other reasonable views on this complex issue. How do you, personally, resolve the differences among these opinions? Somewhere in your essay, we would like to see a description of the main issue, a summary of Lopez's view of it, and a comparison and contrast of his view with the views of at least two others.

REFERENCES

Adam, J. (1979). Sequential strategies and the separation of age, cohort, and time-of-measurement contributions to developmental data. *Psychological Bulletin, 85*, 1309-1316.

Aiken, L. R. (1998). *Human development in adulthood.* New York: Plenum Press.

Astin, A. W. (1987). Assessment, value-added, and educational excellence. In D. F. Halpern (Ed.), *Student outcomes assessment: What institutions stand to gain* (pp. 89-107). San Francisco: Jossey-Bass.

Astin, A. W. (1997). *What matters in college? Four critical years revisited.* San Francisco: Jossey-Bass.

Banta, T. (1997). Moving assessment forward: Enabling conditions and stumbling blocks. In P. J. Gray & T. W. Banta (Eds.), *The campus-level impact of assessment: Progress, problems, and possibilities* (pp. 79-91). San Francisco: Jossey-Bass.

Basseches, M. A. (1987). *Dialectical thinking and adult development.* Norwood, NJ: Ablex.

Bereiter, C. (1980). Development in writing. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 73-93). Hillsdale, NJ: Lawrence Erlbaum.

Bolinger, D. (1980). *Language—the loaded weapon: Use and abuse of language today.* London: Longman.

Brewer, J., & Hunter, A. (1989). *Multimethod research: A synthesis of styles.* Newbury Park, CA: Sage.

Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975). *The development of writing abilities* (pp. 11-18). London: Macmillan.

Charney, D. (1996). Empiricism is not a four-letter word. *College Composition and Communication, 47*, 567-593.

Chiseri-Strater, E. (1991). *Academic literacies: The public and private discourse of university students.* Portsmouth, NH: Heinemann.

Christensen, F. (1968). The problem of defining a mature style. *English Journal, 57*, 572-579.

Christensen, F., & Christensen, B. (1967). *Notes toward a new rhetoric: Nine essays for teachers.* New York: Harper and Row.

Cirillo, L., & Wapner, S. (Eds.). (1986). *Value presuppositions in theories of human development.* Hillsdale, NJ: Lawrence Erlbaum.

Clark, F. (1987). *Studying the longitudinal study: Do impromptu essays show changes in critical thinking of the span of college?* East Lansing, MI: Educational Data Retrieval Service. (ERIC ED 293 120)

Curry, W., & Hager, E. (1987). Assessing general education: Trenton State College. In D. F. Halpern (Ed.), *Student outcomes assessment: What institutions stand to gain* (pp. 57-65). San Francisco: Jossey-Bass.

Dixon, E. (1970). *Indexes of syntactic maturity.* East Lansing, MI: Educational Data Retrieval Service. (ERIC ED 091 748)

Durst, R. K. (1984). The development of analytic writing. In A. Applebee (Ed.), *Contexts for learning to write* (pp. 79-102). Norwood, NJ: Ablex.

Eckhardt, C. D., & Stewart, D. H. (1979). Towards a functional taxonomy of composition. *College Composition and Communication, 30*, 338-342.

Evans, R., & Ballance, C. (1980). A comparison of sentence connective recall by two populations of readers. *Journal of Educational Research, 73*, 324-329.

Evans, R. V. (1979). The relationship between reading and writing of syntactic structures. *Research in the Teaching of English, 13*, 129-135.

Faigley, L. (1979). The influence of generative rhetoric on the syntactic maturity and writing effectiveness of college freshmen. *Research in the Teaching of Writing, 13,* 197-206.

Faigley, L. (1980). Names in search of a concept: Maturity, fluency, complexity, and growth in written syntax. *College Composition and Communication, 31,* 291-300.

Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication, 32,* 400-414.

Freedman, A., & Pringle, I. (1980). Writing in the college years: Some indices of growth. *College Composition and Communication, 31,* 311-324.

Freedman, S. W. (1979). How characteristics of student essays influence teacher grades. *Journal of Educational Psychology, 71,* 328-338.

Freedman, S. W. (1984). The registers of student and professional expository writing: Influences on teachers' responses. In R. Beach & L. S. Bridwell (Eds.), *New directions in composition research* (pp. 334-347). New York: Guilford Press.

Gebhard, A. O. (1978). Writing quality and syntax: A transformational analysis of three prose samples. *Research in the Teaching of English, 12,* 211-232.

Gee, J. P. (1994). *Social linguistics and literacies: Ideology in discourses* (2nd ed.). London: Taylor & Francis.

Graham, J. G. (1987). *A comparison of the writing of college freshmen and college seniors with a focus on indications of cognitive development.* Unpublished dissertation, University of Maryland.

Grobe, C. (1981). Syntactic maturity, mechanics, and vocabulary as predictors of quality ratings. *Research in the Teaching of English, 15,* 75-85.

Haas, C. (1994). Learning to read biology: One student's rhetorical development in college. *Written Communication, 11,* 43-84.

Hagen, L. B. (1971). An analysis of transitional devices in student writing. *Research in the Teaching of English, 5,* 190-201.

Hake, R. L., & Williams, J. M. (1981). Style and its consequences: Do as I do, not as I say. *College English, 5,* 433-451.

Halliday, M.A.K., & Hasan, R. (1976). *Cohesion in English.* London: Longman.

Harman, H. H. (1976). *Modern factor analysis* (2nd ed.). Chicago: University of Chicago Press.

Haswell, R. (1989). Textual research and coherence: Findings, intuition, application. *College English, 51,* 305-319.

Haswell, R. H. (1986). *Change in undergraduate and post-graduate writing performance (Part I): Quantified findings.* Pullman, WA: Educational Data Retrieval Service. (ERIC ED 269 780)

Haswell, R. H. (1990). *Change in undergraduate and post-graduate writing (Part II): Problems in interpretation.* Pullman, WA: Educational Data Retrieval Service. (ERIC ED 323 537)

Haswell, R. H. (1991). *Gaining ground in college writing: Tales of development and interpretation.* Dallas, TX: Southern Methodist University Press.

Haswell, R. H. (1995). *The WSU Portfolio Examination: First findings.* Pullman: Washington State University Office of Writing Assessment.

Haswell, R. H. (1998). Multiple inquiry in the validation of writing tests. *Assessing Writing, 5,* 89-109.

Haswell, R. H. (1999). Rubrics, prototypes, and exemplars: Categorization and systems of writing placement. *Assessing Writing, 5,* 231-268.

Hays, J. N. (1982). *The effects of audience considerations upon the revisions of a group of basic writers and more competent junior and senior writers.* East Lansing, MI: Educational Data Retrieval Service. (ERIC ED 204 802)

Henning, G., & Davison, F. (1987). *Scalar analysis of composition ratings.* Los Angeles: Educational Data Retrieval Service. (ERIC ED 287 285)

Herrington, A., & Curtis, M. (2000). *Persons in process: Four stories of writing and personal development in college.* Urbana, IL: National Council of Teachers of English.

Hiatt, M. (1977). *The way women write.* New York: Teachers College Press.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels.* Champaign, IL: National Council of Teachers of English.

Hunt, K. W. (1970). Syntactic maturity in schoolchildren and adults. *Monographs of the Society for Research in Child Development, 35* (National Council of Teachers of English Research Report, No. 3).

Hurtgen, J. R. (1997). Assessment of general learning: State University of New York at Fredonia. In P. J. Gray & T. W. Banta (Eds.), *The campus-level impact of assessment: Progress, problems, and possibilities* (pp. 59-69). San Francisco: Jossey-Bass.

Johnstone, B. (1996). *The linguistic individual: Self-expression in language and linguistics.* New York: Oxford University Press.

Kaufer, D. S., Hayes, J. R., & Flower, L. (1986). Composing written sentences. *Research in the Teaching of English, 20,* 121-141.

Keil, F. C. (1987). Conceptual development and category structure. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 175-200). Cambridge, UK: Cambridge University Press.

Kerek, A., Daiker, D. A., & Morenberg, M. (1980). Sentence combining and college composition [Monograph Supplement 1-V51]. *Perceptual and Motor Skills, 51,* 1059-1157.

Kitzhaber, A. R. (1963). *Themes, theories, and therapy: The teaching of writing in college.* New York: McGraw-Hill.

Light, R. J. (1993). *Explorations with students and faculty about teaching, learning, and student life* (Harvard Assessment Seminars, 2nd Rep.). Cambridge, MA: Harvard University Graduate School of Education and Kennedy School of Government.

Loban, W. (1976). *Language development: Kindergarten through grade twelve.* Urbana, IL: National Council of Teachers of English.

Loehlin, J. C. (1992). *Latent variable models: An introduction to factor, path, and structural analysis.* Hillsdale, NJ: Lawrence Erlbaum.

Maimon, E. P., & Nodine, B. F. (1979). Measuring syntactic growth: Errors and expectations in sentence-combining practice with college freshmen. *Research in the Teaching of English, 12,* 233-244.

McAdams, D. P. (1993). *The stories we live by: Personal myths and the making of the self.* New York: Morrow.

Mellon, J. C. (1979). Issues in the theory and practice of sentence combining: A twenty-year perspective. In D. A. Daiker, A. Kerek, & M. Morenberg (Eds.), *Sentence combining and the teaching of writing: Selected papers from the Miami University Conference* (pp. 1-38). Akron, OH: University of Akron Press.

Newberry, R. A. (1967). Objective indices in the assessment of essays. *British Journal of Educational Psychology, 37,* 403-405.

North, S. M. (1987). *The making of knowledge in composition: Portrait of an emerging field.* Upper Montclair, NJ: Boynton/Cook.

Ochs, E. (1993). Indexing gender. In B. D. Miller (Ed.), *Sex and gender hierarchies* (pp. 146-169). Cambridge, UK: Cambridge University Press.

Perkins, K., & Brutte, S. R. (1990). Writing: A holistic or atomistic entity? *Journal of Basic Writing, 9,* 75-84.

Raforth, B. A., & Combs, W. (1983). Syntactic complexity and readers' perception of an author's credibility. *Research in the Teaching of English, 17,* 165-169.

Raforth, B. A., & Rubin, D. L. (1984). The impact of content and mechanics on judgments of writing quality. *Written Communication, 1,* 446-458.

Rest, J. R. (1979). *Development in judging moral issues.* Minneapolis: University of Minnesota Press.

Richards, F. A., & Commons, M. L. (1984). Systematic, metasystematic, and cross-paradigmatic reasoning: A case for stages of reasoning beyond formal operations. In M. L. Commons, F. A. Richards, & C. Armon (Eds.), *Beyond formal operations: Late adolescent and adult cognitive development* (pp. 92-119). New York: Praeger.

Scharton, M. (1989). Models of competence: Response to a scenario writing assignment. *Research in the Teaching of English, 23,* 163-180.

Schön, D. A. (1983). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions.* San Francisco: Jossey-Bass.

Schumacher, G. M., Klare, G. R., Cronin, F. G., & Moses, J. D. (1984). Cognitive activities of beginning and advanced college writers: A pausal analysis. *Research in the Teaching of English, 18,* 169-187.

Selzer, J., & Crowley, S. (in press). *Rhetorical bodies: Toward a material rhetoric.* Madison: University of Wisconsin Press.

Spack, R. (1997). The acquisition of academic literacy in a second language: A longitudinal case study. *Written Communication, 14,* 3-62.

Sternglass, M. S. (1993). Writing development as seen through longitudinal research: A case study exemplar. *Written Communication, 10,* 235-261.

Sternglass, M. S. (1997). *Time to know them: A longitudinal study of writing and learning at the college level.* Mahwah, NJ: Lawrence Erlbaum.

Stewart, M. F. (1978). Syntactic maturity from high school to university: A first look. *Research in the Teaching of English, 12,* 37-46.

Sullivan, F. J., Jr. (1997). Calling writers' bluffs: The social production of writing ability in university placement testing. *Assessing Writing, 4,* 53-81.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words.* New York: Teacher's College Press.

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

Vandament, W. E. (1987). A state university perspective on student outcomes assessment. In D. F. Halpern (Ed.), *Student outcomes assessment: What institutions stand to gain* (pp. 25-31). San Francisco: Jossey-Bass.

Watson, C. (1983). Syntactic change: Writing development and rhetorical context. In M. Martlew (Ed.), *The psychology of written language: Developmental and educational perspectives* (pp. 127-139). London: Wiley.

White, E. M. (1989). *Developing successful college writing programs.* San Francisco: Jossey-Bass.

Whitla, D. K. (1980). *Value added and other related matters. Report for the National Commission on Excellence in Education.* East Lansing, MI: Educational Data Retrieval Service. (ERIC ED 228 245)

Williams, J. M. (1979). Defining complexity. *College English, 40,* 595-609.

Williams, J. M. (1981). *Style: Ten lessons in clarity & grace.* Dallas, TX: Scott, Foresman.

Winer, B. J. (1962). *Statistical principles in experimental design.* New York: McGraw-Hill.

Witte, S. E., & Davis, A. S. (1982). The stability of t-unit length in the written discourse of college freshmen: A second study. *Research in the Teaching of English, 16*, 71-84.

Witte, S. P., & Faigley, L. (1981). *A comparison of analytic and synthetic approaches to the teaching of college writing. TWRG Research Report No. 1. Department of English, University of Texas*. East Lansing, MI: Educational Data Retrieval Service. (ERIC ED 209 677)

Witte, S. P., & Faigley, L. (1983). *Evaluating college writing programs*. Carbondale: Southern Illinois University Press.

Witte, S. P., & Flach, J. F. (1994). Notes toward an assessment of advanced ability to communicate. *Assessing Writing, 1*, 207-246.

Wolcott, W. (1994). A longitudinal study of six developmental students' performance in reading and writing. *Journal of Basic Writing, 13*, 14-40.

Wolk, A. (1970). The relative importance of the final free modifier: A quantitative study. *Research in the Teaching of English, 4*, 59-68.

Richard H. Haswell is Haas Professor of English at Texas A&M University–Corpus Christi, where he teaches 1st-year composition, evaluation and diagnosis of writing, professional publishing, and post–World War II American poetry. Comp Tales: An Introduction to College Composition through Its Stories, *with Min-Zhan Lu, was published by Addison Wesley Longman (2000). A monograph on the writing assessment program at Washington State University, with William Condon, Fiona Glade, Lisa Johnson-Shull, Diane Kelly-Riley, Richard Law, Galen Leonhardy, Susan McLeod, Jennie Nelson, and Susan Wyche, is forthcoming from Ablex.*