DOCUMENT RESUME

ED 247 572 CS 208 410

AUTHOR Neuner, Jerome L.
TITLE Cohesion in Teaching and Evaluation: Problems and Implications.
PUB DATE [83]
NOTE 32p.; Tables may be marginally legible.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cohesion (Written Composition); College Freshmen; Comparative Analysis; Connected Discourse; *Discourse Analysis; Higher Education; *Sentence Structure; Syntax; *Writing Evaluation; *Writing Instruction; *Writing Research
IDENTIFIERS *Text Structure

ABSTRACT

Good and poor explanatory essays of 40 college freshmen were analyzed for 18 cohesive ties and chains to determine the appropriateness of the cohesion system for teaching and evaluating writing. The questions that were specifically addressed were, (1) How do writers use the cohesive resources of the language? and (2) How is cohesion related to teachers' perceptions of writing quality? The analysis revealed that the density of ties and length of chains increased disproportionately to the length of essays. A review of individual specimen essays suggested that greater variety and maturity of lexical choice characterized the good essays. Poor essays had frequent pseudochains--long strings of common high-frequency words bearing very little semantic import. Most good and poor essays had a dominant chain connecting several paragraphs. The findings suggest that the cohesion system lacks content and domain selection validity to be appropriate as an evaluation scheme. The system could be used in instruction by a teacher at the point of responding and suggesting revisions, but not as the central emphasis of instruction. (HOD)

COHESION IN TEACHING AND EVALUATION:

PROBLEMS AND IMPLICATIONS

Jerome L. Neuner
Instructor, Academic Development
Canisius College
Buffalo, New York 14208
716-883-7000 ext. 207, 208

Abstract

This research studied cohesive ties and chains in the good and poor explanatory essays of 40 college freshmen and questioned the appropriateness of the cohesion system for teaching and evaluating writing. Of 18 cohesion variables studied, only 1 showed significant difference between the good and poor essays. In addition, the length of cohesive ties from coherer to precursor did not distinguish the good from poor writing, but the length of cohesive chains when corrected for length of essay was a strong discriminator of good and poor essays. A study of good and poor essays by length (long or short) indicated that the density of ties and length of chains increased disproportionately to the length of essays. A review of individual specimen essays suggested that greater variety and maturity of lexical choice characterized the good essays. Poor essays also have frequent pseudo chains, long strings of common high-frequency words bearing very little semantic import. Most good and poor essays had a dominant chain connecting several paragraphs together. The methodology of the research included means' tests and ANOVA for statistical comparisons as well as the examination of cases. The phi coefficient was used to measure the interrater reliability of cohesion analysis. The findings of the complete study strongly suggested that the cohesion system lacks content and domain selection validity to be appropriate as an evaluation scheme. The system could be used in instruction by a teacher at the point of responding and suggesting revisions, but not as the central emphasis of instruction. The terminology of the cohesion system is valuable in that it supplements the terminology of traditional and transformational grammars.

Cohesion in Teaching and Evaluation:
Problems and Implications

The publication of Halliday and Hasan's Cohesion in English (1976) has engendered a large body of research by English educators, much of which appears to have as its goal estimating the usefulness of the cohesion system in evaluating and teaching composition. The research reported here will suggest that researchers have jumped too quickly from theory and description of cohesion to applying it as an emphasis in instruction or a method of evaluation. In other words, cohesion research stands in danger of repeating the sequence of events that occurred with transformational sentence combining: from theoretical background (Hunt, 1965; Mellon, 1969) to application in teaching and evaluation (O'Hare, 1973) to a virtual cottage industry of sentence combining books that professed to work wizardry (Strong, 1976) on a student's composing ability. Throughout this history some more skeptical voices were heard (Marzano, 1976 ; Shaughnessy, 1976) but on the whole the movement to use sentence combining in teaching could genuinely have been called a bandwagon; and the variables associated with syntatic density (mean t-unit length, mean length of clause) became aspects of many evaluation programs at schools and colleges. It took a body of more precise and careful research (Nold and Freedman, 1977; Gebhard, 1978; Stewart and Grobe, 1979) to demonstrate that teacher's evaluations of student writing were not so closely tied to syntatic maturity and that, by implication, a pedagogy emphasizing sentence combining could not deliver all the comprehensive outcomes first promised for it.

Thus there is an urgency to do precise and basic research in the cohesion system before it becomes a part of the received but untested wisdom of teaching and evaluation. Already some early reviewers of Halliday and Hasan's system (Holloway, 1981; Witte and Faigley, 1981) have given cautiously favorable estimates. And more recent textbooks (Williams, 1981) have begun to use the terminology of cohesion in their discussions of transitions and sentence connection. A detailed analysis of cohesion has not yet been used in a large scale writing assessment, but reports from the 1980 National Assessment of Educational Progress indicate that readers can be directed to use a simplified cohesion rubric in assessing compositions (Odell, 1981, 122-123). It is much too early to say that another bandwagon is forming, but it does not seem to be true that many teachers and researchers feel an attraction to cohesion and its apparent power to describe textual relationships. English teachers and researchers seem to practice frequently a kind of iron law of novelty: if some new insight from linguistics or psychology appears on the horizon, use it in teaching and evaluation until more exhaustive research questions its usefulness. Should not the opposite occur? Meticulous and careful questioning should precede the widespread application of linguistic theory or systems to teaching or evaluating writing. Researchers must not adopt a new terminology for teaching or new variables for research until they have evidence that the terminology is not simply a new jargon and that the variables genuinely distinguish good from poor writing.

How do writers use the cohesive resources of the language, and how is cohesion related to their teachers' perceptions of writing quality? These two questions are the focus of this research. Their answers have broad

implications for both pedagogy and evaluation. The review of research described below emphasizes those studies directed either explicitly or implicitly to evaluation or teaching. The cohesion system itself is, of course, fully described by Halliday and Hasan. All summaries are inevitably reductive and imperfect, but all researchers to date have simplified the system for their research uses.

Review of Research

Eiler (1979) took an exhaustive look at the writing of 15 ninth-grade honor students and discovered that various kinds of lexical cohesion seemed to be the best indicator of the students' response to literature and that reference cohesion was the primary evidence of ability to sustain a self-sufficient ("endophoric," in the terminology of Halliday and Hasan) text without appeal to the non-textual ("exophoric") environment. Hartnett (1980) tried to teach the cohesion system to basic writers at a Texas college and then used counts of different kinds of ties as a criteria for evaluation of the essays. She had mixed results, with no significant differences found for teacher, treatment, or mode of writing for the experimentals over the controls. The teacher x treatment interaction was significant, but in general the correlation of holistic score with number of types of cohesive ties was quite small, only .2076 for all essays.

Cherry and Cooper (1980) studied average and superior writers at grades four, eight, twelve, and college. They introduced some interesting variables such as the average distance of ties (by number of intervening t-units between coherer and precursor) and the relative dispersion of ties in the first, second, and third thirds of essays. Their basic conclusion was that as writers mature they seemed to rely more on lexis and less on reference and

conjunction. (Substitution and ellipsis were rare.) The proportion of ties that were lexical went up from 56% to 59% to 63% to 68% as students ascended across the four grades studied. Pritchard (1980) studied the good and poor compositions of 44 eleventh graders and discovered that the average use or frequency of total lexical or grammatical ties did not distinguish the good from the poor essays. On the other hand, she found that the notion of "cohesive problem" does have some empirical validity since passages marked by her readers as "problem sections" varied from other sections by their proportional use of ties. Pritchard concluded that counts of cohesive ties are not measures of their effectiveness and her conclusions are especially convincing. She used statistical transformations to stabilize the distributions and then repeated her tests using three different sets of variables: (1) average number of ties per 100 words, (2) frequency of ties per 100 words, and (3) frequency of ties per t-unit. No single type of tie was found to be a significant discriminator of good and poor essays in all three schemes, and the ANOVA test for all types of ties was also nonsignificant in each case.

Witte and Faigley (1981) studied five good and five poor freshman essays by using a simplified list of ties with frequency counts (ties per 100 t-units) and relative percentages as their variables. Their findings were similar to those for Cherry and Cooper's twelfth graders: about two-thirds of ties were lexical, and good essays seemed to have greater density of all types of ties. (No statistical tests were performed to compare good and poor writing.) They concluded that cohesion appeared to be an important property of writing, but no evidence suggested that large or small numbers of ties in themselves effect writing quality.

Most recently Tierney and Mosenthal (1983) studied 24 essays ranked by teachers for general coherence and divided into two different topics (a biographical sketch and a thematic essay) under conditions of the writers' familiarity or unfamiliarity with the subject (determined by whether the writers had seen a filmstrip). They found that cohesion varied by topic, with biographical sketches having a somewhat larger proportion of reference ties and thematic essays a larger proportion of lexical ties. But cohesive patterning did not predict rankings on general coherence. There was a familiarity x text topic interaction when looking at coherence rankings but not when looking at cohesive proportions. The main point was that familiarity, topic, and coherence did not seem related to the specifically linguisitic aspect of texts detailing the use of lexical and reference ties. Cohesion was pervasive in all texts but causally unrelated to coherence. Tierney and Mosenthal used some interesting new varibles such as the ratios of pronouns and lexical ties to total ties (P+L/T) and temporal conjunctives to total conjunctive ties (TC/T).

Several generalizations arise from this body of research. The research varies from highly exacting to more casual in quality, and the studies using the more precise techniques (inferential statistics, reliability coefficients, greater number of cases, data transformations) are more cautious in recommending cohesion as a teaching or testing method. Another theme that appears is the search for cohesion variables that appear to distinguish good from poor writing, the writing of younger from that of older students, and the different modes or purposes of writing. These matters are not yet fully resolved. The research reported here will suggest some further considerations in the choice of variables and reliability coefficients and

offer some more explicit remarks on the usefulness of cohesion.

Method

The present research was an effort to replicate with greater precision a comparison of cohesive devices in good and poor freshman essays written on a single explanatory topic (i.e., research similar to that of Pritchard, Cherry and Cooper, and Witte and Faigley) but to expand and improve upon it by examining new variables: the interrater reliability of cohesion analysis, a more complete list of types of ties, the relative distances between coherers and precursors, the mean length of cohesive chains, the dispersion of ties within texts, and the effects of length of essay. In addition, since nearly all researchers had argued for the need to examine cohesion non-statistically in individual essays, an analysis of specimen essays was also performed. The logic of this comparison was quite simple. If the cohesion system is to be useful as an evaluation method, it ought to show great variation across levels of writing quality. Choosing the very best and very poorest essays in a large representative sample provides the greatest extremes of quality.

Essays were collected from a sample of over 600 written by college freshmen at a summer orientation and testing session. The conditions were carefully controlled: each student received the topic assignment during check-in and had two hours of relatively unorganized time including a lunch to think about or discuss the assignment with peers. All essays were written in fifty-minute sessions during the afternoon in proctored classrooms. The students themselves represented a wide range of abilities and aptitudes. Their SAT scores (combined verbal and math) ranged from 650 to 1400 and their high school averages ranged from 72.4 to 98.2. The entire sample of essays

was read by a panel of twelve college professors who have had two to five years of experience with holistic scoring and who have demonstrated interrater reliability exceeding .90 in their use of a 1 (low) to 4 (high) holistic scale. The 20 good essays were selected randomly from among those that had received the highest holistic score (4) from two holistic readers and the 20 poor essays from those that had received the lowest score (1) from two readers.

The cohesion analysis was then performed by the researcher and two other English teachers after careful instruction and practice on essays from the original orientation sample. One of the teachers had participated in a previous cohesion study and was expert in using the system. The second had no experience and required approximately eight hours of instruction and practice before he could recognize satisfactorily the types of ties considered in this research. The analysts worked with carefully written directions and had extensive practice in recognizing the types of ties studied in this research. A strict interpretation of cohesive relations was respected at all times: (a) each coherer had to have an identifiable and literal precursor in a prior t-unit; (b) the list of common coherers was extensive; (c) erroneous or ambiguous references were not counted; and (d) cases of multiple cohesion were counted as distinct individual ties. For a complete description of these directions as well as examples of the coding and analysis protocols, see Neuner 1983, 134-139.

Results

Reliability

In this study the phi coefficient (ϕ) (Kurtz and Mayo, 1979, 346-355) was used to measure the interrater reliability two-by-two for the three cohesion

analysts. Phi is used when the raw data are dichotomous choices (yes or no, male or female) and was computed for this research by going through the readers' coding sheets for several sample essays and tallying word-by-word their agreement or disagreement on the cohesive status of every word in the text. This word-by-word analysis of reliability is much more rigorous than various rank order coefficients such as Kendall's τ used by Pritchard in her study, which used the ranking of most to least frequent types of cohesive ties in an essay. The three phi coefficients for the analysts in this study were .839 (readers 1 and 2), .887 (readers 1 and 3), and .828 (readers 2 and 3). These coefficients are much higher than are usually found for most types of essay scoring procedures and approach the .90 reliability usually required of standardized achievement tests. Many researchers have neglected the interrater reliability of cohesion analysis even though most recognize the number of "judgment calls" frequently required in using the system. The figures reported here suggest that the system can be used reliably if careful instruction and practice are provided. Readers who were not already experienced English teachers and essay graders would undoubtedly require many more hours of study and practice.

Proportions of Cohesive Ties

Table 1 illustrates the raw numbers and percentages of the various types of cohesive ties as well as the results of 18 separate t-tests to compare the means of each type of tie across the good and poor essays. The table provides the totals and percentages, but the statistical tests were performed on data transformed by a square root function recommended by Pritchard who cites Snedecor and Cochran (1957, 325-329). This analysis reconfirms the findings of many researchers to date: a simple counting of ties does not

appear to distinguish good from poor writing at this level. The various percentages of ties do not vary radically from good to poor essays. Comparative reference (line 4 Table 1) is the only type of cohesive tie that shows a significant difference between good and poor. Several other analyses were performed to verify this important conclusion, including a comparison of the average words per tie in each essay and a comparison of the average

Table 1

Cohesive Ties in Good and Poor Essays

type of cohesive tie	Good Essays N = 20		Poor Essays N = 20		t	probability value
	number	%	number	%		
1. pronouns	130	9.9	89	12.9	.89	.379
2. demonstratives	77	5.9	35	5.1	1.19	.242
3. definite articles	41	3.1	13	1.9	.69	.495
4. comparatives	20	1.5	3	.0	2.55	.015*
5. total reference	268	20.5	140	20.4	.48	.637
6. substitution/ellipsis	26	2.0	8	1.2	.94	.353
7. additive conjunctions	44	3.4	22	3.2	.57	.571
8. adversative conjunctions	50	3.8	29	4.2	.34	.738
9. causal conjunctions	11	.1	8	1.2	.37	.712
10. temporal conjunction	36	2.7	13	1.9	.98	.334
11. continuatives	1	.0	6	.9	.88	.385
12. total conjunction	142	10.1	78	11.4	.72	.479
13. same item	546	41.7	293	42.7	.24	.814
14. synonym/hyponym	91	6.9	58	8.5	.17	.866
15. superordinate	32	2.4	19	2.8	1.47	.151
16. general item	12	.1	6	.9	.27	.792
17. collocation	193	14.7	84	12.2	1.44	.157
18. total lexical	874	66.7	460	67.1	1.02	.313
Totals	1310	100.0	686	100.0		

degrees of freedom = 38 for all comparisons

* $p < .05$

Total words: good essays 7811 poor essays 4265

number of lexical ties using data transformed by an arc sine function (Snedecor and Cochran, 1967, 325-329). These comparisons also failed to show statistically significant differences between the two groups of essays: $t(38) = 1.37$, and $t(38) = .18$; p nonsignificant in each case.

Relative Cohesive Distances

Halliday and Hasan (1976, 330-331) have a scheme for describing the form of cohesive ties relative to the number of intervening sentences between coherer and precursor. The scheme includes simple immediate ties, mediate ties, remote ties, and mediated-remote ties. In this research it was decided to dispense with the system in favor of a simpler counting of the intervening t-units between precursors and coherers even if the immediate precursor was not the original source of primary meaning. A relationship longer than an individual precursor-coherer pair was defined as a cohesive chain: a series of references, collocations, reiterations, synonyms, or superordinates all semantically related to one another. For each essay the total distances for all ties were summed up and then divided by the total number of ties in that essay to provide the average length in t-units of the cohesive ties. Then that average length was divided by the number of t-units in the essay to provide an average relative distance from coherer to precursor. For example, one essay was 26 t-units long, had 57 cohesive ties, with a total distance of 119 t-units between the various coherers and precursors. Its average relative distance was: $119 \div 57 \div 26 = .080$. The same strategy was used to determine the average relative length of cohesive chains. The total distance of the three or four longest chains was divided by the number of chains, which was then divided by the number of t-units in the essay. For example, one essay had a total distance of 48 t-units for 4 cohesive chains and was 23

t-units long. The average relative length of those chains was: 48 ÷ 4 ÷ 23 = .522. The purpose of using these relative figures was to correct for the different lengths of essays. Obviously an essay 30 t-units long is likely to have longer chains and ties than an essay 15 t-units long. So a relative rather than an absolute average length must be computed in order to compare essays on this variable.

Table 2 illustrates a comparison of the relative average distances of ties and chains in the good and poor essays and in all essays. It must be remembered that these are relative figures and not the true average lengths of ties and chains. This analysis suggests that good and poor essays are not distinguished by the distances of individual ties if length of essay has been accounted for. However, the distance of chains does discriminate good from poor essays even if length of essay is accounted for. This is another way of saying that good essays seem to be more intensely about their subjects than poor essays are, regardless of which essay is longer. A word that Halliday

Table 2

Relative Average Distances
of Cohesive Ties and Chains

Variable	All essays N = 40	Good essays N = 20	Poor essays N = 20	t	probability value
Average relative distance, coherer to precursor	.099	.103	.095	.60	.552
Average relative length of chains	.586	.647	.525	2.86	.007 **

degrees of freedom = 38

p. < .01

and Hasan like to use about this relationship is "texture." This research seems to suggest that texture resides more in cohesive chains than in individual precursor-coherer ties.

Dispersion of Ties

Cherry and Cooper (1980) had suggested that good essays would have their cohesive ties more evenly spaced throughout the writing while poorer essays would have their ties more cumulated toward the end. To test this hypothesis each essay was divided into thirds by t-unit count and the number ot ties in each third was tallied. Table 3 reports the percentages of total ties in the first, second, and third thirds of both good and poor essays.

Table 3

Percentages of Cohesive Ties in First, Second, and Third Thirds of Good and Poor Essays

Section of Essay	Good Essays N = 20	Poor Essays N = 20	Difference	F	Probability Value
First Third	26.7%	24.8%	1.9%	.452	.506
Second Third	32.6%	36.0%	3.4%	1.769	.192
Third Third	40.6%	39.2%	1.4%	.238	.589

Degrees of freedom: between groups 1
within groups 38
total 39

Table 3 implies clearly that both good and poor essays are roughly cumulative in that the concentration of ties increases regularly from the first to second to third thirds of essays. However, in each third the difference between good and poor essays does not approach statistical significance.

A good argument can be made on theoretical grounds that nearly all types of texts in every mode are likely to be generally cumulative with respect to cohesive ties. As a text evolves, more and more words come into existence, and this fact makes it more likely that later words will be the coherers for earlier precursors. In addition, as a text evolves, greater opportunities for multiple cohesive ties occur in the later words. If for example a student wrote an essay about the different sports teams he or she had played on and wrote a concluding sentence such as "I certainly enjoyed all these sports," then the word sports would be a multiple coherer (a superordinate) for the names of all the individual sports mentioned earlier in the essay. The same principle seems to be true for nearly every text regardless of subject. The only discourse for which this would not be true would be nontexts such as lists and inventories.

Length of Essay

No researcher to date has attempted to estimate the effects that length or brevity of text have on the proportions or distributions of cohesive ties or chains. Pritchard (1980) for example carefully chose texts of about the same length, 250 to 300 words, to rule out length of text as a confounding variable. Texts may have characteristics of recursiveness, iterativeness, and unevenness which make it an error to assume that long and short texts are simply macro and micro versions of each other with respect to cohesion. To explore this matter more carefully, the 6 longest and 6 shortest essays from

both the good and poor groups were analyzed for proportions and dispersion of ties and the distances of ties and chains. These data along with several measures of essay length are reported for the good and poor essays in Table 4. It should be understood that these measures were not transformed by square root or arc sine functions nor adjusted for the relative length of essay. The purpose of these comparisons was precisely to look at untransformed data to observe the effects of essay length on the various kinds of ties and distances.

Table 4

Cohesive Ties and Distances in Long and Short Good Essays and Poor Essays

	Good Essays		Poor Essays	
Variable	Long (N = 6)	Short (N = 6)	Long (N = 6)	Short (N = 6)
Average Length				
Words	521.6	293.7	309.7	105.2
t-units	39.2	21.2	24.8	6.5
Words/t-units	13.3	13.8	12.5	16.1
Average Number of ties				
per 100 words	18.1	15.4	16.4	11.3
per t-unit	2.4	2.1	2.0	1.8
Percentages of Ties				
Reference	20.2%	19.9%	24.3%	23.9%
Substitution/Ellipsis	3.3%	1.8%	1.3%	1.4%
Conjunction	9.7%	9.6%	13.5%	8.5%
Lexical	66.7%	68.8%	61.2%	66.2%
First Third	28.5%	23.9%	31.9%	16.9%
Second Third	34.3%	33.8%	33.9%	39.4%
Third Third	37.1%	42.3%	34.2%	43.7%
Mean Cohesive Distances (in t-units)				
Coherer to Precursor	4.00	2.24	2.38	.77
Length of Chains	27.75	12.50	14.74	3.25

Table 4 tends to suggest that the various percentages of types of ties do not differ dramatically from the long to the short essays in either good or poor categories. The percentages of reference, substitution/ellipsis, conjunction and lexis are strikingly similar in the good and poor essays and also similar to the findings for the entire group of 40 essays studied in this research.

Regarding the dispersion of ties, some differences appear between the long and short essays, especially in the essays of poor quality. However, an important artifact of the cohesion system tends to imbalance these percentages: in the coding of cohesive ties the first t-unit of a text has no cohesive ties because there is no prior text in which precursors can be found. (I have excluded from this research the extremely rare instances of cataphoric cohesion, ties in which the precursor item comes after rather than before the coherer.) This removal of the first t-unit from those potentially available to contain coherers has an inordinate effect on the shortest of essays. For example, if an essay is only 6 t-units long there are 2 t-units in each third of the essay. But the first third really has only 1 t-unit available for coherers since no coherer can exist in the very first t-unit of the text. On the other hand, removing the first t-unit from a much longer text has a proportionally smaller effect on the number of ties in the first third of that text. This removal of 1 t-unit from each essay accounts for almost all the variation apparent in the dispersion of ties in the first third of essays across long and short texts. If the first t-unit is excluded from consideration, the long and short essays demonstrate the same general cumulativeness discovered in the entire sample of good and poor essays (Table 3).

The mean distance figures in Table 4 are more interesting. In each category the mean distances from coherers to precursors are almost exactly proportional to the length in words of the essays in that category. In other words, as the lengths of essays vary, the average distance from precursor to coherer seems to vary in the same proportion. However, this even proportionality is not true for the mean length of chains. The good long essays are 77.6% longer in words and 84.9% longer in t-units than the good short essays (521.6 to 293.7 in words; 39.2 to 21.2 in t-units) but the chains in good long essays are 122% longer than they are in good short essays. The poor long and short essays reflect similar differences. Poor long essays are 193% longer in words and 282% longer in t-units than the poor short essays (309.7 to 105.4 in words; 24.8 to 6.5 in t-units), but the chains in poor long essays are 353% longer than they are in poor short essays. In other words, as essays become longer, the length of their cohesive chains becomes longer at an even greater rate.

The figures for density of ties per t-unit and per 100 words show a small effect for t-units and a larger effect for words. Ties per t-units vary from 1.8 for the shortest poor essays to 2.4 in the longest good essays. But ties per 100 words appear to vary more substantially from shortest to longest essays: 11.3 for poor short essays (105.2 words); 15.4 for good short essays (293.7); 16.4 for poor long essays (309.7 words); and 18.1 for good long essays (521.6 words). A greater appreciation of these values can be attained by considering lexical ties, which are the most frequent type in every case: in long good essays, one of every 8.26 words is a lexical coherer; in long poor essays, one in every 9.98 words; in short good essays, one in every 9.42 words; in short good essays, only one in every 14.67 words. These figures

make a strong argument that lexical cohesion flourishes where many words are available to set up the reverberations of synonyms, hyponyms, collocations, superordinates and reiterations. Conversely, lexical cohesion languishes where many fewer words are available.

In summary, this admittedly exploratory glance into length of essay has suggested that texture does and does not differ according to length of text. Clearly the various percentages of types of ties and the dispersion of ties (taking into account the removal of the first t-unit as a source of ties) do not appear to change substantially as length varies. However, the density of cohesive ties per 100 words and the length of cohesive chains do appear to vary substantially as length changes. Both are important differences in their own right and also important because, as every teacher knows, in any given classroom writing situation the better essays tend also to be longer essays. It may be the case that, for example, the differences between good and poor essays discovered by Witte and Faigley (1981) were really only differences related to length of essay and that the same differences would have been discovered if all the essays were poor but some were much longer than others. Future research should attempt to take into account these effects of text lengths whenever investigating essays of widely varying length.

No *t* or *F* values were computed for Table 4 because the number of cases was very small, and so these findings must be considered a first glance rather than a definitive study of texture by length of text. Also to be resisted is the desire to universalize these results. Studies of other types of writing and discourse may reveal other effects.

Specimen Essays

A random selection of 10 essays, 5 good and 5 poor, was examined word-for-word and the items in the major cohesive chains tabulated in columns across the t-units and paragraphs in rows. This technique, an adaptation of a method suggested by Halliday and Hasan (1976, 16), is illustrated by the following good essay and its tabulation in Table 5. The subscript numbers indicate the t-unit count and the italicized words indicate items in the dominant chain.

My rather short life has been plagued with unsolicited advice.$_1$ It is one of life's ironies that only unwanted advice is given.$_2$ When one really needs another's opinion one is immediately told to think things out alone.$_3$

One flint-like nugget of advice that I once received was to continue my upward climb out of the darkness of ignorance by going on in mathematics.$_4$ My following this advice was a mixed blessing indeed.$_5$ I learned that math could be fun, when I wandered upon the correct answers.$_6$ I also learned that where numbers are concerned there is no light at the end of the tunnel,$_7$ there is always more to learn.$_8$ Numbers and I regard each other warily.$_9$ In fact, I have great respect for the power they have over me; the power to frustrate and the power to make situations clearer.$_{10}$

In my pursuit of higher mathematics I discovered letters accompanying numbers, then letters standing alone.$_{11}$ This interested me as letters are my forte.$_{12}$ I realized the piece of advice I received could even prove helpful, (although I cannot recall ever encountering an x on the street).$_{13}$ It has been said that a math is the language God wrote the universe with.$_{14}$ This means math is a powerful, universal, skill.$_{15}$ Although numbers seem foreign to my nature I have realized they are just another form of communication.$_{16}$ Through mathematics other worlds, and even our world, can be explored to a greater extent.$_{17}$

My unsolicited advice to you is to realize that numbers have been arbitrarily chosen by man.$_{18}$ It is possible to form your own number system.$_{19}$ The system in use today works well,$_{20}$ but it does the frustrated math student's heart good to understand this point.$_{21}$ Numbers are only as powerful as you make them.$_{22}$

Table 5

Dominant and Minor Chains in a Good Essay

Paragraph	t-unit number	Dominant Chain	Minor Chain 1	Minor Chain 2	Minor Chain 3
1	1	-	advice	-	-
	2	-	advice	-	-
	3	-	opinion?	-	-
2	4	mathematics	advice	-	darkness ignorance
	5	-	advice	-	
	6	math	-	-	
	7	numbers	-	-	light
	8		-	-	learn?
	9	numbers	-	-	
	10	they	-	-	power power frustrate power clearer
3	11	mathematics numbers	-	letters letters	-
	12	-	-	letters	forte?
	13	-	advice	an x	-
	14	math	-	language wrote?	-
	15	math	-	-	powerful
	16	numbers they	-	(form of) communication	-
	17	mathematics	-	-	-
4	18	numbers	advice	-	-
	19	number system	-	-	-
	20	system	-	-	-
	21	math	-	-	frustrated
	22	numbers them	-	-	powerful

? signifies an item that could be challenged for its place in the chain.

A tabulation such as this provides at a glance a visualization of the length of chains (in t-units), the number of items in chains, the amount of iterativeness and variety in lexical items, the degree to which chains are confined to or extended beyond paragraphs, and the places where chains intersect in individual t-units. The tabulation also suggests some of the decisions a researcher must make on whether certain words are lexically related and belong in the same chain. Some good examples are in Minor Chain 2: are power and light collocations? Is to frustrate a true opposite of to make...clearer? Does forte belong in the chain? These questions might cause genuine arguments among different people, and no simple method to resolve them yet exists.

For the purposes of comparisons, below is a poor essay and Table 8 tabulates its chains.

As a graduating senior I would like to pass on a word of advise to all of you.$_1$ As I was entering my first year at our beloved school, a graduating senior of that year told me, "Always listen to your parents.$_2$ As you get into your latter teens you start to "break away" from your parents, because "your old enough to make your own decisions"$_3$ and "your friends are allowed to."$_4$ But they have already lived through everything you are going through,$_5$ so you should listen to them.$_6$ Most of your parents have gone out drinking or have gone somewhere they weren't supposed to go.$_7$ But now when you do it, they ground you.$_8$ You probably figure, that they are just trying to be mean or show you who's boss.$_9$ But most parents (I can't say all) are not anything like that.$_{10}$ The reason they punish you, is because they love you and don't want anything to happen to you.$_{11}$ They always (well, at least most of the time) have a good reason for not letting you do some of the things you do.$_{12}$ What if when they were in High School one of their

friends went to a bar, got drunk, drove home, or at least tried to, and got in an accident, possibly even killing someone.[13] You have to understand your parents and communicate with them, talking things over and learning to understand the reasons for doing things they do." Through my four years I tried to communicate.[15] And I found that I became a lot closer to my parents and that there were less arguments.[16] I hope you take this word of advice and let it help you through the years.[17]

Table 6
Dominant, Minor, and Pseudo Chains in a Poor Essay

Paragraph	t-unit number	Dominant Term	Minor Chain 1	Minor Chain 2	Pseudo Chain
1	1		pass on		
	2	parents	listen		
	3	parents	-	-	-
	4	-	-	-	-
	5	they	-	-	everything
	6	them	listen	-	-
	7	parents they	-	-	-
	8	they	-	-	-
	9	they	-	-	-
	10	parents	-	-	anything
	11	they they	- -	reason -	anything -
	12	they	-	reason	things
	13	they their	-	-	-
	14	parents them they	communicate talking	reasons -	things things
	15	-	communicate	-	-
	16	parents	-	-	-
	17	-	-	-	-

Such tabulated lists are, of course, entirely asyntactic and offer no insight into the hierarchy of ideas or the transitions between them. They provide merely a skeleton of lexical items in a loose systems network. Despite these shortcomings, a close examination of the tabulations suggests some patterns that identify or distinguish them.

1. Good and poor essays alike have what Markels (1981) has called a dominant term, a word or phrase more or less continually present either directly (by reiterations) or inferentially (by synonyms, collocations, superordinates, and pronoun references). This dominant chain provides a reservoir of associations to which the writer returns frequently for elaboration and predication as the discourse proceeds.

2. Poor essays occasionally have a dominant chain that simply overwhelms the essay with the reiteration of its topic and pronouns for the topic. The poor essay above on listening to parents illustrates this pattern. The term parents and its pronouns appear 18 times in 17 t-units and there are only 15 terms in the other 3 chains combined. Good essays, on the other hand, have a dominant chain that constitutes a smaller proportion of the total items. As an example, the good essay on mathematics has 18 items in its dominant chain, but the three minor chains contain a total of 27 items.

3. Good essays have greater variety (i.e., more different words) and maturity (i.e., words of lower frequency in the language as a whole and greater explicitness) in their chains. This feature can be observed in the dominant chain of the good essay (Table 5) which has mathematics, math, numbers and system along with pronouns. By comparison the poor essay (Table 6) has only parents and pronouns in its dominat chain.

4. Poor writers have pseudo chains, non-cohesive strands of words such as thing, do, way, be, know, and have. These words collocate with virtually every word in the language and therefore bear little semantic import or explicitness. The pseudo chain in Table 6 illustrates clearly.

5. Good essays have more real chains, and poor essays have fewer meaningful minor chains, the weakest of which comprise only 3 or 4 items. Minor Chain 2 in Table 6 is an example of this feature.

6. Chains can be related to paragraphs in several ways: chains may tie together several paragraphs (and in fact the whole essay) or may be almost completely confined in a single paragraph (Minor Chain 2 in Table 5). In other cases, chain items may be heavily concentrated in one paragraph and then reoccur more sparingly in others.

Discussion

It may be argued that these observations are not in any sense new. They are the kinds of remarks that teachers make about student writing, only now they are couched in the language of cohesion. For example, Shaughnessy (1977), in her classic study of errors among basic college writers, has a chapter on vocabulary that notices many of the same findings observed in the lexical chains tabulated for this research: inadequate synonymies; extensive reiteration without variety; repetition of high-frequency, low-significance words; inappropriate word choice. Is cohesion, therefore, just a new jargon or has anything useful been learned? Halliday and Hasan (1976, 327-328) clearly distinguish between what a text means and how it attains that meaning. The study of cohesion in a text does not add to what the text means, but can help in understanding how and why it attains that meaning. In addition, notions

such as chain, tie, coherer, precursor, context, and texture all suggest valuable truths about discourse: that connectedness is at the core of meaning; that no discourse is without relationship to its environment no matter how remote; that meaning is resolved not in a single word or phrase but through longer semantic structures. Readers often speak of following a text, of writers' leading readers to a conclusion, of stories flowing to their endings, of articles running too long. Thus the terminology of cohesion is in keeping with common descriptions of the meaning-making process. Teachers must speak to students about writing, and they must therefore share some lexicon of terms for their talk. The traditional terminologies of the parts of speech or structural parts of sentences all shared the weakness of implying that texts are made from static, unitary, and unmoving segments like bricks or stones. The terminology of transformational and generative grammar suggested notions of depth and surface, with sentences having a history, the story of transformations as sentences come upward from deep to surface structure (Myers, 1981, 10). The terminology of cohesion, on the other hand, suggests an idea of interrelatedness in networks or sytems and a notion of flow or movement. No one terminology should supplant all the others. As Young, Becker, and Pike argued in their famous (1971) text, Rhetoric: Discovery and Change, entities should be observed and interpreted as particle, wave, and field in order to be fully understood. Cohesion has given teachers a terminology to discuss the field aspects of text-making.

Implications for Evaluation and Teaching

The major limitation of the cohesion system is that it describes only one-third of a complete theory of language, namely the textual function or

text-forming component of the linguistic system. The other two functions of language in Halliday and Hasan's (1976, 26-27) system are the ideational and interpersonal. The ideational is concerned with content, with the function that language has of being about something. The interpersonal is concerned with social, expressive, and conative functions; with expressing attitudes, judgments, role relationships and motives for speaking or writing. The cohesion system does not and cannot capture or describe these ideational or interpersonal qualities of language. Consequently, cohesion should not be the exclusive or central emphasis of either a pedagogy or an evaluation method for writing.

In the language of test-makers and evaluators, the cohesion system lacks content validity or domain selection validity (Popham, 1975, 120, 156-159). In other words, measuring cohesion is not measuring enough of all the things that should be measured if one wishes to accurately evaluate a language act. Classroom teachers even more than researchers will feel constrained if they attempt to rely heavily on estimates or counts of cohesive ties in evaluating student writing, because such an emphasis may detract attention from the quality of ideas, the sense of audience, the development of purpose, and the creation of persona, all of which should be more central to evaluation. Cohesion is also on the borderline of usability, another important quality of evaluation systems. The full list of all types of ties in Halliday and Hasan's system represents 176 categories and subcategories. Only 18 plus the distance values were reported in this research, but that represents more categories than most teachers would wish to deal with for purposes of routine classroom evaluation.

Ann Ruggles Gere (1980) argues persuasively that an evaluation method based on all three of Halliday and Hasan's language functions would be superior to such current methods as analytic scales, primary trait scoring, general impression scales, and holistic scoring. Such a method would represent the first theoretically sound evaluation system. But the attempt to use cohesion as the primary emphasis for evaluation would be doomed to failure. Cohesion, except for certain characteristics of cohesive chains and distances, is simply not sufficiently related to writing quality. In addition, the research reported here finds some aspects of cohesion sensitive to length of essay, and this could introduce confusion into the evaluation of writings that vary greatly in length.

Regarding pedagogy, only one controlled and systematic attempt to teach cohesion as the emphasis in instruction and then evaluate essays by counting types of ties has been published. Hartnett (1980) had mixed results, with no significant differences being found for teacher, treatment, or mode of writing for the experimentals over the controls and a relatively low correlation (.2076) between the number of different cohesive ties (Hartnett's basic variable) and holistic score.

The research reported here suggests that a better use of the cohesion system would be in a teacher's responding to a full piece of writing and in suggesting revisions. For example, pointing out reiterations in a text could help a student understand excesses or insufficiences of redundancy. By circling items in a cohesive chain, a teacher could help explain his or her perceptions of the too rapid or too slow pace and development of the student's ideas. Pointing out the distances between precursor and coherer items could help the student appreciate the teacher's difficulty in

following the thread of meaning if too much space intervenes between terms in a cohesive relation. As Holloway (1981) has suggested, this kind of instruction could become not just a new jargon but an invitation for students themselves to ask new and higher-order questions about their writing, including: "What is the old information I need to present so that I can tie this idea to what has gone before?" and "Have I reflected the connections of thought in my thesis statement implicitly or explicitly in the devices I have used for cohesion in the body of my paper?" Questions such as these combine the insights of the cohesion scheme with the concerns for ideational and interpersonal functions necessary to make the text coherent as well as cohesive. Coherence implies that the ideational and interpersonal functions of language must be respected as first principles.

In conclusion, this research clearly suggests that the cohesion scheme not be over-prompted as a testing or evaluation method or as a central emphasis in the teaching of writing. Counting up cohesive ties in general did not distinguish the good from the poor writing examined in this study. And cohesion is surely too narrow an aspect of language to have as the center of teaching writing.

References

Cherry, Roger & Cooper, Charles. Cohesive ties and discourse structure: a study of average and superior texts at four grade levels. Unpublished manuscript. Department of Learning and Instruction, State University of New York at Buffalo. 1980.

Eiler, Mary Ann. Meaning and choice in writing about literature: a study of cohesion in the expository texts of ninth graders. Unpublished doctoral disseration, Illinois Institute of Technology, 1979.

Gebhard, Ann O. Writing quality and syntax: a transformational analysis of three prose samples. Research in the Teaching of English, 1978, 12, 221-233.

Gere, Ann Ruggles. Written composition: toward a theory of evaluation. College English, 1980, 42, 44-58.

Halliday, M.A.K. & Hasan, R. Cohesion in English. London: Longmans, 1976.

Hartnett, Dale. Semantic grammars: how they can help us teach writing. College Composition and Communication, 1981, 32, 205-218.

Hunt, Kellog. Grammatical structures written at three grade levels. Champaign, Ill.: National Council of Teachers of English, 1965.

Kurtz, Albert & Mayo, Samuel. Statistical methods in education and psychology. New York: Springer-Verlag, 1979.

Markels, Robin Bell. Cohesion patterns in English expository paragraphs. Unpublished doctoral dissertation, Ohio State University, 1981.

Marzano, Robert. The sentence combining myth. English Journal, 1976 65, 57-59.

Mellon, John. Transformational sentence combining. Champaign, Ill.: National Council of Teachers of English, 1969.

Myers, Miles. Approaches to the teaching of composition, in M. Myers and J. Gray (Eds.), Theory and practice in the teaching of composition. Champaign, Ill.: National Council of Teachers of English, 1983.

Neuner, J.L. A study of cohesion in the good and poor essays of college freshman. Unpublished doctoral dissertation, State University of New York at Buffalo, 1983.

Nold, E.W. & Freedman, S.W. An analysis of readers' responses to essays. Research in the Teaching of English, 1977, 11, 164-174.

Odell, Lee. Defining and assessing competence in writing in C. Cooper (Ed.), The nature and measure of competence in English. Urbana, Ill.: National Council of Teachers of English, 1981.

O'Hare, Frank. Sentence combining: improving student writing without formal grammar instruction. Urbana, Ill.: National Council of Teachers of English, 1973.

Popham, W. James. Educational evaluation. Englewood Cliffs, N.J.: Prentice Hall, 1975.

Pritchard, Ruie Jane. A study of cohesion devices in the good and poor compositions of eleventh graders. Unpublished doctoral dissertation, University of Missouri, 1980.

Shaughnessy, Mina. Errors and expectations. New York: Oxford University Press, 1977.

Snedecor, George & Cochran, William. Statistical methods. 6th ed. Ames, Iowa: Iowa State University Press, 1967.

Stewart, Murray & Grobe, Gary. Syntactic maturity, mechanics of writing, and teachers' quality ratings. Research in the Teaching of English, 1979, 13, 207-215.

Strong, William. Sentence combining: back to the basics - and beyond. English Journal, 1976, 65, 60-64.

Tierney, Robert & Mosenthal, James. Cohesion and textual coherence. Research in the Teaching of English, 1983, 17, 215-229.

Williams, Joseph. Style: ten lessons in clarity and grace. Glenview, Ill.: Scott Foresman, 1981.