This study is the third in a series of reports of the Reading-to-Write Project, a collaborative study of students' cognitive processes at one critical point of entry into academic performance. This part of the study examines the problem that teachers have in judging whether textual signals that students use to indicate a persuasive analysis of source material are or are not enacted in their discussions. Subjects, 69 students enrolled in a freshman composition course (36 in the control group and 33 in the experimental group) had their essays graded by a group of teachers. Fifty-seven students later handed in revised essays. In discussions some essays came in for special scrutiny because the raters encountered various difficulties when they applied their initial taxonomy of organizing plans to the freshman papers. This led to the development of an elaborated taxonomy of task representations. (Two figures are included; and the Reading-to-Write study list of references, instructions to judges of essays, and sample essays are attached.) (RS)
PROMISES OF COHERENCE, WEAK CONTENT, AND STRONG ORGANIZATION:
AN ANALYSIS OF THE STUDENT TEXTS
(Reading-to-Write Report No. 3)

Margaret J. Kantz

May, 1989

This Report will appear as a chapter in Reading-to-Write: Exploring a Cognitive and Social Process, by Linda Flower, Victoria Stein, John Ackerman, Margaret J. Kantz, Kathleen McCormick, and Wayne C. Peck, to be published by Oxford University Press. An overview of the Study to which this Report refers can be found in CSW Technical Report No. 21, Studying Cognition in Context: Introduction to the Study.
CENTER FOR THE STUDY OF WRITING

Director
Sarah Warshauer Freedman
University of California, Berkeley

Co-Directors
Linda Flower
Carnegie Mellon University
J.R. Hayes
Carnegie Mellon University
James Gray
University of California, Berkeley

Academic Coordinator
Sandra Schecter
University of California, Berkeley

Editor
Melanie Sperling
University of California, Berkeley

Publication Review Board
Chair
Kay Losey Fraser
University of California, Berkeley

Assistant Chairs
Carol Heller
University of California, Berkeley
Victoria Stein and Lorraine Higgins
Carnegie Mellon University

Advisors
Charles Fillmore
University of California, Berkeley
Jill H. Larkin
Carnegie Mellon University

Millie Almy, University of California, Berkeley
Carla Asher, Herbert H. Lehman College of the City University of New York
Nancie Atwell, Boothbay Region Elementary School, Boothbay Harbor, ME
Robert de Beaugrande, University of Florida
Ruby Bernstein, Northgate High School, Walnut Creek, CA
Lois Bird, Whole Language Consultant
Wayne Booth, University of Chicago
Robert Caffee, Stanford University
Michael Cole, University of California, San Diego
Collette Daiute, Harvard University
John Daly, University of Texas, Austin
Peter Elbow, State University of New York, Stony Brook
JoAnne T. Eresh, Writing and Speaking Center, Pittsburgh, PA
Donald Graves, University of New Hampshire
Robert Gundlach, Northwestern University
James Hahn, Fairfield High School, Fairfield, CA
Julie Jensen, University of Texas, Austin
Andrea Lunsford, Ohio State University
Marian M. Mohr, Fairfax County Public Schools, Fairfax County, VA
Lee Odell, Rensselaer Polytechnic Institute
Charles Read, University of Wisconsin
Victor Rentel, Ohio State University
Michael W. Stubbs, University of London
Deborah Tannen, Georgetown University
Henry Trueba, University of California, Santa Barbara
Gordon Wells, Ontario Institute for Studies in Education
Preface to the Reading-to-Write Reports

So I'm just gonna--I don't care, I'm just going to interpret them the only way I can interpret them. . . . Let's just put what the authors agreed on. Authors agree -- We'll just -- If at least two of them concur, we'll say they agree. Authors in general agree that . . . But then they don't agree -- There's nothing you can say about this . . .

Can I leave it at that . . . Oh give me a break, I don't know what I'm doing. I'm only a freshman. I have no idea what to do.

Darlene, a first-semester freshman

Darlene's college assignment asked for synthesis and interpretation. The paper she turned in--a short, simplistic review of material from her sources--failed to meet her own expectations and her readers'. And yet, a chance to look at the process behind this unsophisticated product revealed serious thinking, a complicated, if confused, decision process, and a trail of unused abilities and discarded ideas--an active encounter with academic discourse that her teacher would never see.

The study presented here takes an unusually comprehensive look at one critical point of entry into academic performance. It shows a group of freshmen in the transition into the academic discourse of college, looking at the ways in which they interpret and negotiate an assignment that calls for reading to write. On such tasks, students are reading in order to create a text of their own, trying to integrate information from sources with ideas of their own, and attempting to do so under the guidance of a purpose they must themselves create. Because these reading-to-write tasks ask students to integrate reading, writing, and rhetorical purpose, they open a door to critical literacy. Yet this same interaction often makes reading-to-write a difficult process for students to learn and to manage.

In order to get a rounded picture of cognition in this academic context, the study looks at the thinking processes of these students from a number of perspectives, drawing on think-aloud protocols of students writing and revising, on interviews with and self-analyses by the students, and on comparisons of teachers' and students' perceptions of texts the students wrote. It attempts to place these observations within a broader contextual analysis of the situation as students saw it and the social and cultural assumptions about schooling they brought with them.

What this study revealed were some radical differences in how individual students represent an academic writing task to themselves--differences which teachers might interpret as a simple indication of a student's ability rather than a student's interpretation of the task. The students were often unaware that such alternative representations existed or that they might hold such significance. Some images of the task, for instance, such as those dominated by the goals of comprehension, summary, and simple response, offered little or no place for critical response, original synthesis, or interpretation for a rhetorical purpose.

The reading-to-write task students imagined for themselves also had a direct effect on performance: it affected the goals they set, the strategies they used, and the
ways they solved problems during composing. And it led to differences in teachers' evaluations of the texts—although, this study suggested, these evaluations may confuse the conventions of organization (e.g., use of topic sentences) with the writer's control of ideas. When students began to examine their options and attempt the more demanding task of interpreting for a purpose, certain students, whom we called the Intenders, showed important changes in their writing and thinking process. These changes, however, were not evident in the text and nor apparent to teachers. Finally, this study showed how students' images of the task were rooted in the students' histories, the context of schooling, and cultural assumptions about writing which they brought to college.

It is not surprising to find that some of the images students bring with them are at odds with the expectations they encounter at a university. However, when the expectations for "college-level" discourse are presented in oblique and indirect ways, the transition students face may be a masked transition. That is, the task has changed, but for a number of reasons, the magnitude and real nature of this change may not be apparent to students, even as they fail to meet the university's expectations.

One of the key implications of this study is that reading-to-write is a task with more faces and a process with more demands than we have realized. We see students thinking hard and doing smart things, even when they misgauge their goals or their written text fails to meet certain standards. This close survey of the cognitive and social landscape of reading-to-write in a college class gives one added respect for the students in this transition and for the complexity and sophistication of the "freshman" task as they face it.

The Reading-to-Write Project was carried out as a collaborative effort at the Center for the Study of Writing, at Carnegie Mellon. We designed the study to create a range of alternative perspectives on the process of reading-to-write and on the way cognition is shaped by the social context of school. The following technical reports present the design and collaborative history of the study; analyses of the cognitive processes we observed, of the texts, and of students' perceptions of both; and a set of conclusions, from different theoretical perspectives, on how students manage this entry into academic discourse:


Reading-to-write is an act of critical literacy central to much of academic discourse. This project, divided into an Exploratory Study and a Teaching Study, examines the cognitive processes of reading-to-write as they are embedded in the social context of a college course.

Reading-to-Write Report 2. (CSW Tech. Report 6) The Role of Task Representation in Reading-to-Write. Linda Flower

The different ways in which students represented a "standard" reading-to-write task to themselves led to marked differences in students' goals and strategies as well as their organizing plans. This raised questions about the costs and benefits of these alternative representations and about students' metacognitive control of their own reading and writing processes.
Promises of Coherence, Weak Content, and Strong Organization: An Analysis of the Student Texts. Margaret J. Kantz

Analysis of students' Organizing Plans (including free response, summary, review and comment, synthesis, and interpretation for a rhetorical purpose) also revealed a hybrid plan in which certain coherence conventions gave the promise of synthesis while the paper's substance reflected a simpler review and comment strategy. Both students and teachers, it appeared, may sometimes confuse coherence strategies (for text) with knowledge transformation strategies (for content).

Students' Self-Analyses and Judges' Perceptions: Where Do They Agree? John Ackerman

Any writing assignment is a negotiation between a teacher's expectations and a student's representation of the task. Students' Self-Analysis Checklists showed a strong shift in perception for students in the experimental training condition, but a tellingly low agreement with judges' perceptions of the texts.

Exploring the Cognition of Reading-to-Write. Victoria Stein.

A comparison of the protocols of 36 students showed differences in ways students monitored their comprehension, elaborated, structured the reading and planned their texts. A study of these patterns of cognition and case studies of selected students revealed both some successful and some problematic strategies students brought to this reading-to-write task.

Elaboration: Using What You Know. Victoria Stein

The process of elaboration allowed students to use prior knowledge not only for comprehension and critical thinking, but also for structuring and planning their papers. However, much of this valuable thinking failed to be transferred into students' papers.

The Effects of Prompts Upon Revision: A Glimpse of the Gap between Planning and Performance. Wayne C. Peck

Students who were introduced to the options of task representation and prompted to attempt the difficult task of "interpreting for a purpose of one's own" on revision were far more likely to change their organizing plan than students prompted merely to revise to "make the text better." However, the protocols also revealed a significant group of students we called "Intenders" who, for various reasons, made plans they were unable to translate into text.
One context for writing is the student's history of schooling including high school assignments and essays. Based on protocols, texts, and interviews, this report describes a set of "initial reading strategies" nearly every freshman used to begin the task—strategies that appear to reflect their training in summarization and recitation of information. From this limited and often unexamined starting point, students then had to construct a solution path which either clung to, modified, or rejected this a-rhetorical initial approach to reading and writing.

By setting reading-to-write in a broad cultural context we explore some of the cultural imperatives that might underlie particular cognitive acts. Protocols and interviews suggest that three culturally-based attitudes played a role in this task: the desire for closure, a belief in objectivity, and a refusal to write about perceived contradictions.

Entering an academic discourse community is both a cognitive and social process guided by strategic knowledge, that is, by the goals writers set based on their reading of the context, by the strategies they invoke, and by their awareness of both these processes. As students move from a process based on comprehension and response to a more fully rhetorical, constructive process, they must embed old strategies within new goals, new readings of the rhetorical situation. However, for both social and cognitive reasons, this process of negotiation and change that academic discourse communities expect may not be apparent to many students for whom this becomes a confusing and tacit transition.

One important implication of this entire study is that students themselves should come into the act of examining their own reading and writing processes and becoming more aware of cognitive and cultural implications of their choices. This set of classroom approaches, written by teachers collaborating on a Reading-to-Write course that grew out of this project, introduces students to ways of exploring their assumptions and alternative ways of represent aspects of the task.

**Acknowledgments**

Our heartfelt thanks go to our colleagues John R. Hayes, Karen A. Schriver, Nancy Spivey, Tom Huckin, Christina Haas, Lorraine Higgins, Stuart Greene, Tim Flower, Stephen Witte, Mike Rose, Gerald Rutledge, and Kathy R. Meinzer.
PROMISES OF COHERENCE, WEAK CONTENT, AND STRONG ORGANIZATION:
AN ANALYSIS OF THE STUDENT TEXTS

By

Margaret J. Kantz
Texas Christian University

It is a fact of life that students interpret writing tasks in different ways. These task representations matter because they affect the written products. Differences in task representation lead to marked differences in papers that may in turn determine whether individual papers fulfill a teacher's expectations and how they are evaluated. Sometimes, however, students appear to think about a task in one way but do it in another way, producing papers that send conflicting textual signals which confuse the teachers who read them. In the study being reported in this report, teachers who operated on the basis of seemingly reasonable expectations for structure and coherence found themselves temporarily unable to agree about the differences in particular papers—and, hence, unable to agree on how the writers had interpreted the assignment. In the papers that caused this difficulty, for example, the introductions might announce a plan to argue or discuss information, but the bodies of the papers would use coherence-building strategies that matched a quite different task interpretation, such as summarizing the information. We identified half a dozen strategies for announcing or building coherence. In some essays, readers felt that the discussion of the Time Management material matched the strategy; in others—the papers that they found so difficult—the discussion did not match the strategy. We feel that the seemingly inconsistent task representations embodied in these papers—a phenomenon not clearly predicted by the literature—are worth discussing because they may shed light on a common problem that teachers have in reading student papers, namely, judging whether the textual signals that students use to indicate a persuasive analyses of source material are or are not enacted in their discussions.

This report first describes the ways that readers saw the structures in a set of freshman papers, giving the results of the Teaching Study described in Report 2. The readers' analysis was motivated by the question, "Given the alternative organizing plans this task elicited in the Exploratory Study, what choices would the students in the Teaching Study make?" The second half of the report discusses the problem that the judges encountered when they applied the initial taxonomy of organizing plans, developed in the Exploratory Study, to these freshman papers. This problem, involving mismatches between content and structural cues, led to the development of a somewhat elaborated taxonomy of task representations. This elaborated taxonomy (the new sub-categories are indented) is summarized here and was used to code the essays, as reported below in Figures 1 and 2 (see Report 2 for definitions of the five major categories; instructions to judges are found in Appendix I).

**Summary**
states gists or selected ideas from the source text

**Review and Comment**
combines a summary or selective review of material from the source text with commentary or additions by the writer
**Isolated Main Point or Conclusion**

a subcategory of review and comment containing a statement which is signalled as main point or a conclusion, but which does not appear to control the structure or selection of content in the body of the text

**Frame**

a subcategory of review and comment containing a statement which frames the review with a vague, highly general, or obvious introductory statement

**Free Response to the Topic**

discusses the topic with little reference to information from the source texts

**Synthesis**

organizes a discussion (which draws on source materials) around a unique (non-obvious) controlling concept

**Interpretation for a Purpose of One's Own**

organizes a discussion (which draws on source materials) around a unique and apparent rhetorical purpose (beyond the purpose to summarize, comment, synthesize etc. See Appendix II for examples.)

Figures 1 and 2 show the organizing plans that students used on their original and revised essays, as the readers saw those plans using the elaborated taxonomy (See Appendix III for interrater reliability). Out of the 72 students originally enrolled in the four sections of the freshman course, three did not hand in either an original or a revised essay and were dropped from the study, bringing the number of subjects observed to 69. Of these 69 students, all of whom handed in an original essay, 36 were in the control group and 33 were in the experimental group. Fifty-seven of these students (82%) later handed in revised versions of their essays. Of these 57 students, 32 were in the control group and 25 were in the experimental group.
Distribution of original essays across coding categories. N=69. Black columns indicate control group. Bricked columns indicate experimental group.

Figure 1: Choices students made about organizing plan for original essays.

Figure 2. Choices students make about organizing plan for revised essays.
Figure 1 shows the choices students made on the original assignment; Figure 2 shows their choices when they revised the essay. The figures show clear differences in the types of plans used in the original and revised essays; in particular, students in both the experimental and control groups moved from using review-and-comment plans to using controlling concepts that actually did control their discussions. This result suggests that in this study the act of revision apparently stimulated the students in both the Experimental and Control groups to rethink their interpretation of the writing task. Students in the Experimental section were alerted to their own choices and asked to attempt the task of interpreting for a purpose. They heard a lecture on the options we had observed to date (see Report 1, Appendix II), worked through a personal checklist (see Report 4, Appendix 1), and revised their discussions using directions to choose a rhetorical purpose. The Control group, randomly selected within each section of the course, were asked to revise their text "to make it better." They did not hear the lecture or use a personal checklist until after completing the revision. Although the original essays in both groups are distributed across all the categories, 68% are clumped in the knowledge-telling group that includes Summary plus the various forms of Review and Comment; the Review-and-Comment-with-Frame category alone accounts for nearly a quarter of the papers. Among the revised essays of both groups, more students wrote essays using a controlling concept.

The experimental manipulation did appear to have an effect, however. More students in the Experimental group wrote revised papers with rhetorical interpretations of the source than did students in the Control group. A Chi Square test was used to determine whether the number of essays coded as transforming knowledge (i.e. controlling concept or interpretation) differed significantly between the original and revised groups. The obtained result was significant at the .01 level ($X^2 = 7.6, df=1$). To understand the nature of and the rationale for these shifts, however, we need a more detailed analysis of what individual writers did. An analysis of the revision process that students brought to this task and a detailed comparison of the two groups will be found in Report 6.

Rationale for the Two Taxonomies

Before we can understand why certain essays seemed to us to embody inconsistent representations of the assigned task, we need to examine the essays, especially those that caused the readers' difficulties, in order to learn what characteristics produced the confusion. But before these essays are discussed, it is necessary to understand how the essays were identified as deserving of special scrutiny.

We began our study by developing a five-element taxonomy (given in Report 2, Figure 2) which describes ways that students might approach our writing task. This taxonomy is not a theoretical description of academic writing tasks in general; it was designed in the context of our task. Nor is it intended for use in a testing situation. Performance testing requires a highly explicit writing prompt (task) that produces clean, reliably coded responses to be used in discriminating levels of ability. When writing is used for assessment, as Ruth and Murphy have shown, a prompt that is open to multiple interpretations can (and they argue, often does) lead to disastrous consequences for some students (1984). In a testing situation one wants to design a prompt that will not only have a uniform interpretation and cue a specific response (that can be described by a primary trait rubric, for instance), one also wants to create enough difficulty to "weed out" weak writers or to discriminate between a good and average performance (Ruth and Murphy, 1984, Smith et al., 1985). In this setting,
differences in interpretation of the task are defined as "response errors," i.e., as failure to design a good prompt.

By contrast, we wanted to study a situation that invited multiple task interpretations in order to observe a range of possible interpretations. Our prompt was thus designed to invite a range of responses rather than to specify a particular response. Using this prompt, we expected to study the process of task representation. We thus needed 1) an open-ended prompt from which the writers had to construct their own plans; 2) the context of a class, so that the web of expectations associated with academic writing was implicit but too conditional or too complex to be spelled out in the assignment; and 3) a taxonomy that described the students' texts without evaluating relative levels of quality. We therefore developed the first taxonomy, a tool that named some of the most distinctive (to us) options that writers took on this task and created a continuum of approaches going from least manipulative (summaries) to most manipulative (interpretations for a persuasive rhetorical purpose). We planned to use the taxonomy as a way to talk to the students about their choices of task interpretations.

In designing this taxonomy, of course, we faced the problem of how to avoid deciding in advance what we would see, i.e., of distorting our understanding of the data by prescribing its interpretation. To avoid this error, we designed the initial taxonomy from three kinds of information. First, we used the organizing plans that seemed most apparent and most distinctive from one another, based on our reading of texts from the three pilot classes of junior-senior undergraduates, M.A., and Ph.D. students. Secondly, we brought some preconceptions about common text types that teachers encounter (summary, synthesis, free response). Third, we brought two theoretical questions that we wanted the textual analysis to comment on: "How do students integrate (or do they integrate) ideas from a source text with their own ideas?" and "To what extent do students carry out any sort of purposeful, rhetorical transformation of their information, as might be reflected in both a synthesis and an interpretation?" Although we expected the resulting five-item taxonomy to oversimplify the possibilities, we decided that it was at least conditionally usable because it met our goals of starting with the categories observed in the pilot work, of maintaining descriptive simplicity, and of achieving some sensitivity to two theoretically interesting features of these texts. It offered a workable tool for analyzing the freshmen texts, or at least the basis for developing such a tool.

The second (final) taxonomy, presented on the first two pages of this report, contains two new categories that describe variations on the review and comment plan: texts marked by an isolated main point or conclusion and texts marked by what we call a frame. The process that led to this elaboration is in some ways typical of any attempt to build a coding scheme. However, the process is also worth reporting because the conflict between the structural cues and the information structure in the texts, which our judges encountered, may reflect a problem that other teachers face in trying to evaluate similar freshmen essays. The problem is this: If the raters read closely for content, they risked failing to recognize attempts to organize the material, and if they read closely for organization, they risked overlooking weak content. Moreover, if the readers read strictly, refusing to admit essays into categories that they might not belong in because of content/structure mismatches, they risked overlooking genuine achievements in the papers. If, however, the readers read generously and admitted the papers into categories to which they did not quite belong, they risked making invalid assumptions about what the writers had tried to do. Before we could make decisions about how the writers had interpreted the task based on their essays, we had to agree on what the essays did. The new taxonomy describes characteristics of student papers
which we suspect occur commonly, in other papers written for other classroom assignments, and which we believe deserve discussion because they confuse readers.

Promises of Coherence: What were Those Essays Doing?

When we created the initial taxonomy, we assumed, incorrectly, that the structural cues that the students had created about how the ideas in their texts were related would indicate whether the writers had actually re-structured or transformed the information from the sources. For instance, one writer might cue readers to expect a synthesis by making an evaluative statement about the material in the first paragraph, e.g., "This material on time management seems very useful." Another writer might cue readers to expect a summary by announcing a plan to discuss three main aspects of the material, e.g., "The most important ways to manage time are planning when to do tasks, managing one's energy level, and creating a good study environment." Readers usually expect the signalled structure of a text (the cues it sets up) to serve the content—to indicate reliably the internal logic of the information presented. This link between content logic and structural conventions is one of the important traits that teachers consider when they evaluate papers (Freedman 1979a, 1979b,1982), and it gives a convenient way to talk about whether students have goals of telling knowledge or goals of transforming knowledge—i.e., paraphrasing sources or using source material to build an original argument, especially an argument that is adapted to a rhetorical purpose. As the thought develops, the structure shifts to accommodate it—so goes the conventional wisdom. When Freedman (1979b), for example, described her study for which she rewrote 32 student essays so that they would be weak or strong in content, organization, sentence structure, and mechanics, she explained that all possible combinations of variables were used except that weak content was never combined with strong organization: "It would have been an exercise in absurdity [our italics] to try to order illogical ideas logically or to order and transition appropriately a group of inherently unrelated ideas" (p. 162).

The essays in the pilot studies appeared to confirm the conventional wisdom. These essays had typically fallen into clear organizational patterns that coincided with the ways the writers had manipulated their information. Summaries, with or without authorial comment, used the order of sources in the time management text. Essays built around an original synthesizing concept typically foregrounded that concept and did not depend on the organization of the original text; instead they used a structure that complemented the controlling concept, such as problem-solution or comparison-contrast. Essays that reinterpreted the assignment for a rhetorical purpose, i.e., talking to a particular audience about Time Management for a particular reason, also used a structure that complemented their controlling idea. Organizing plans as signalled by structural cues, then, seemed a convenient and logical basis for categorizing papers.

Yet when we applied these categories to the undergraduate essays, Freedman's "absurdity" is what we sometimes found: When we tried to place the essays on a continuum of idea-transformation, from those which summarize (tell ideas) to those which transform for an original rhetorical purpose (transform and use ideas), we found a group of essays in the middle of the continuum in which structural signals did not match the treatment of ideas. We found, for instance, apparent summaries sprinkled with comments and evaluations of the source text, arranged in an implied chronological order. We also found texts in which unrelated pieces of source material were arranged as a list and presented with other cues that were apparently meant to suggest a logical sequence. These oddly linked summaries were often framed in the introduction and
conclusion by a larger idea that had an implicit topical relationship to the essay content but that did not in any way control the discussion or presentation of the content. These mismatched textual signals made it very difficult for the raters to decide what approach to the task the writers had been trying to use.

Using the first taxonomy to categorize these texts produced an interrater reliability that was significant but too low to be acceptable. The three raters for the text analysis were experienced teachers and highly motivated members of the research team who after two training sessions and a refinement of the category descriptions had achieved a high degree of agreement on training materials (a random sample of papers written for the time management assignment). The problem, it appeared, was not in the raters, but in the ability of the taxonomy to account for a subset of the texts for which we could not get unanimous judgments. Here is an example of such a text:

**Essay 066:** Extensive research has produced many books on the subject of *Time Management* in professional and educational settings. Experts agree that to work effectively, one must set aside time for a particular purpose. A noted efficiency expert, Alan Lake [sic], says that the secret to success is the pacing and planning of one's "external and internal prime time." External prime time is the time that one extends to helping others. Internal prime time is the time in which one does the best work. It helps to block out this time in advance.

Once this time has been put aside it must be employed well, that is, concentration must be maintained for its duration. This is perhaps the greatest challenge with an even more personal solution than the dilemma of simply setting aside the external or internal prime time. A distraction-free atmosphere is needed for concentration so that one's thoughts may be directed towards a particular project. This means no disturbing sounds, from music to background conversations, should steal your attention. Another nemesis of concentration is mental fatigue. Psychologist, William James states the answer lies in overcoming mental fatigue when it appears. He says that this is done by pushing for continued concentration until a "second wind" is acquired. Another expert, Jean Guitton, believes that when one feels the slightest sign of fatigue it is vital to rest immediately and that one should always approach work with a restful attitude. This is in sharp contrast with James's theory. Concentration is achieved through different means by different individuals. It may take years for one to learn to do so.

Utilization of the time set aside for work requires an answer on the part of every individual. Another problem is how much time to set aside for a particular project. A recent survey of college students revealed that they manage their time in a manner designed to alleviate pressure. They put off long-range projects and finish the easier assignment first in order for them to have fewer projects and therefore fewer distractions. Having one of a number of projects finished seems to take away more pressure than working on many simultaneously.

Time management is much more than simply writing a weekly schedule of work hours. It requires knowing when and where to find quality time, and what to attend to first. This knowledge is gained through experience. Like any other skill, one must practice time management to be successful using it.
This essay was variously scored as a summary, a summary plus comment, and a synthesis. Clearly, it contains summary, some comment, and a conclusion of some kind. One reader argued that the summary develops into an argument that is specifically stated in the final paragraph (Review and Comment); another reader felt that the argument, though not stated clearly until the end, could be said to integrate the discussion throughout (Synthesis). In other words, our readers, who like all good readers were trying to build a coherent text, could (1) find clear cues that this was summary; or (2) they could give more attention to the last paragraph and perceive the selection of statements as an argument/commentary; or (3) they could, by making an even greater constructive effort, see the conclusion, with its broad organizing truism, as a source of coherence for the entire text. Since coherence is something readers construct, the organizing plan they perceive is affected by those structural cues to which they choose to attend and by the significance they give those cues. These texts sent out mixed signals not only about their structure but also about their conceptual depth and the degree to which the information in them was adapted or transformed according to a central plan. Our readers could not decide from the essay how the writer had interpreted the task.

Essay 066, for example, starts off with two sentences that might introduce a controlling concept but which instead lead to a summary of one person's ideas. The second paragraph develops another idea, the effective use of one's time, but the potentially interesting conflicts in the experts' advice are simply presented with the moral, "What works for X may not work for Y." The third idea, deciding how much time to allocate, is stated without reference to the previous ideas and introduces another paragraph of summary. The final paragraph summarizes the essay and adds a moral. Although the essay has an overall topical coherence, and the individual paragraphs cohere, as a whole it cannot be said to develop any original idea about time management, and the three paragraph topic ideas do not treat the topic in a logical order. Indeed, since the essay uses the order of the source text, one could argue that the writer presented its ideas as paragraph topics and just tacked the concluding sentences on to what is basically a summary. Although the pattern of general statements and specific statements is that of the classic five-paragraph theme, in this case the pattern has not produced a logically developed essay. A teacher reading this essay might conclude that it is disorganized, incoherent, and incompetent, and that the writer--far from interpreting the task--simply misunderstood it.

On the other hand, given a combination of good will and constructive effort on the part of the reader, the paper can be read as if it were coherent, as if an intention of overall coherence were implicitly present throughout, finally emerging in the conclusion. "After all," another teacher might say, "paragraph one sets up an approach, based on the process of studying, which controls the order of the discussion. The process moves from setting aside time to using the time, including finding a good place to work and then concentrating on the work. So-and-so is a good student and always tries hard; this may not be her best paper, but I'll give her the benefit of the doubt." The conflicting scores (summary, summary plus comment, and synthesis) given to the essay by the raters indicate that such a sympathetic reading is indeed possible.

We believe that the reasons for the lack of unanimity among our raters are interesting because we believe that other teachers have struggled with mixed messages such as those in essay 066. In such essays, coherence building strategies seem to be used independently of logical relations within the information itself. Such coherence cues include the use of repeated key words, the causal, chronological or additive
phrases which link adjacent paragraphs, and the presence of an idea raised in one paragraph that becomes in some way associated with an idea in the next paragraph. These structure-building cues lead one to expect that the information in the text has been transformed into an argument, a synthesis, or a commentary that will lead to new ideas; yet the substance of the text fails to live up to these promises of coherence and transformation.

Thus, in deciding what should receive priority in their interpretation of the text, the readers were apparently often forced to choose between these two elements—"strong" structure and "weak" content. To provide more information on what the raters were attending to, we asked them to jot comments about their basis for judging these problematic essays. One rater said about a particular essay, "This is a pretty sleazy concept, but there is a concept so this must be a synthesis"; another, however, called a spade a spade: "Never mind the decorations; I know a summary when I see one." The raters were not even consistent in their own readings of such essays: They not only disagreed with each other, but on subsequent readings they sometimes disagreed with themselves, depending, it seemed, on whether they gave priority to structure or to content. (It should be noted here that conflicts between "strict" and "interpretive" readings did not necessarily reflect consistent tendencies to read one way or the other; rather, each rater read sometimes one way and sometimes the other.)

The reasons for this indecision became clear when we sorted the essays according to the categories in conflict (e.g., Summary/Review+Comment, Summary/Synthesis, Summary/Rhetorical Reinterpretation). The largest area of disagreement occurred between the categories of Review-and-Comment and Synthesis. Since we had defined Synthesis as a paper having a controlling concept (a large idea that governed the presentation of material in the essay), this confusion of categories meant that our raters could not agree as to whether our student writers were thinking for themselves (transforming knowledge) or whether they were reproducing their sources (telling knowledge). The essays appeared to mix summary with original thought, and the raters could not agree as to which purpose predominated.

The following excerpt, for example, is not a synthesis, but it is more than a summary or a review with comments. The summary is subordinated to the concept of the various aspects of Time Management, and that concept controls the summary. On the other hand, the controlling idea is so loose that it allows the writer to say almost anything.

**Essay 005:** Something that affects us all, at different times and in different situations, is the problem of time management. Time management is an involved process that can be broken down into many different aspects.

Pacing and planning, according to efficiency expert Alan Lake, [sic] are two very important aspects in time management. [A summary of Lake in follows, then a summary of the rest of the text. Sample paragraph topic sentences include:]

Environment is vital in one's preparation to work efficiently. . . .

Certain strategies are used by students that help reduce pressure caused by up-coming assignments. . . . [The essay concludes:]

17
After taking into account the planning and preparation, the ability to concentrate, a good environment and certain standard strategies, the subject of time management is basically an individual process. Its usage is unique for every individual, whether at home, on the job or at school.

As another example, essay #100 is not a summary or a summary with comments. Although it seems relatively free from the stimulus text, it cannot be considered a free response to the topic. But the list order for presenting ideas might exclude it from being considered a synthesis governed by a controlling concept. One reader might say, "Except for the first sentence, this is a pretty coherent comparison-contrast of two approaches to time management, so I'll call it a synthesis"; another reader might say, "Announcing that 'different views exist' is not my idea of a real concept; this may not be a summary, but it sure isn't a synthesis either! I don't know what it is, and I have no idea what this writer thought he was trying to do."

**Essay 100:** The first and foremost observation about this packet is that it is a poor source of information on time management. Only the first section is about the topic and only some of the others mention it at all. The next observation is the different views on time management given by people in a situation to apply it and those advising people to apply it.

People in a work situation, in this case students, manage time on the assumption that there isn't enough time to do everything. Difficult tasks are scheduled last and work is thought to be better under rushed conditions. Time is precious and never use more than absolutely necessary is the recurring theme among those in the position to actually manage time.

Researchers and others detached from the actual places where time management is employed suggest the opposite theory. Pauk, the researcher, and Lakein, the efficiency expert give a perspective from above and outside the place of work. Their studies suggest scheduling. Assume that you have plenty of time and allow large blocks of time for each individual task. Their strategies seem nowhere near as specific as those of the students.

In general, those involved in time management personally assume there is very little time and they must rush. Those observing and advising say there is lots of time and the slower you go the better. Perhaps if these two views could be integrated better, a full understanding of time management would result.

When the structural cues and promises of coherence were not supported by the substance, some raters still sometimes found themselves led by the promises of such coherence to credit these papers with achieving an integration and transformation of information that others readers were certain they had not achieved. The raters' sense of this conflict was reflected in comments which noted that such papers were "poor" syntheses (a value category our scheme had not provided). Apparently, even experienced readers could be led to conflate superficial coherence (i.e., the lavish use of conventional coherence-building cues) with conceptual integration, even when the promised connections or integrating concept were not fully realized in the text.
This variation may, of course, be what one expects from normal constructive readers. Although we saw this conflict as a striking feature of some freshmen papers, it may have been equally present but unnoticed in the pilot texts, since those papers were not categorized by multiple readers. On the other hand, this tendency to unwittingly conflate the superstructure of coherence with the content structure may occur in other evaluative situations in which teachers must decide whether or not, as readers, to "construct" a coherent student text on the basis of weak cues. For instance, Spivey (1984) found that the discourse syntheses written by less able comprehenders scored lower on "connectivity for the reader" than those by more able college age readers. "Low connectivity" in this case was a measure of how often readers felt that the "flow of reading would be interrupted because of an unusual burden of constructive processing" in which they were forced to reorder, to disambiguate, or to infer plausible connections (p. 16). When Applebee, Durst and Newell, focusing on the text structure rather than the reader, compared passages from published textbooks with comparable analytical essays written by ninth grade students, they found that the students were "more likely to produce a essay with two parallel and unintegrated content structures," i.e., a "mixed" structure that might contain islands of narrative, and thesis/support, question/answer or other organizing patterns (1984, p. 68). Applebee et al. suggest that such "mixed" and unconnected structures are not unusual: "this pattern of relying on familiar structures in the transition to more complicated forms is common as students extend their writing skills" (p. 71).

How do teachers categorize (and evaluate) such papers? In the context of our study, teachers reading such papers may face two dangers: 1) In attending to the underlying summary structure, they might fail to give credit for good intentions or for a conceptual structure that is "present" but buried in a first draft, or 2) by responding to these superficially coherent papers as if they were "genuine" arguments or syntheses or interpretations (as our raters sometimes did), they would be failing to tell students that these more complex organizing plans call for both structural signals and a substantive transformation of information. A response that refused to fill in the gaps might send a more accurate message to students by saying that, for academic readers, coherence must be more than skin deep.

Looking at this experience, at the students' own self-evaluations (see Report 4), and at the confusion that terms such as "synthesis" and "interpretation" caused in the protocols, we came to a disturbing, if tentative, conclusion. The freshman in this study did not seem fully aware of the difference between building coherence through the use of local or superficial cues and building coherence by adapting, restructuring, or even generating information. Their texts promised one thing and did another. Moreover, our judges had difficulty making consistent judgments, depending on which of these powerful but unstated criteria took priority. Neither we nor the teachers nor the students had a ready vocabulary for making this distinction.

Our solution to this problem was, in essence, to provide a vocabulary in the elaborated version of the taxonomy so as to distinguish between coherence-building techniques and a more thorough-going transformation of information. With the help of Thomas Huckin, we first gave these texts a "generous reading," actively looking for the many ways these writers helped the reader construct a coherent text. Such a reading makes one appreciate the large repertoire of such techniques these students possess, even when the techniques are used independently of the knowledge-transforming process that interested us. These promises, with the forms that the essays took, can be described as follows:
Limited promises, e.g., of summarizing, as in "Different experts say different things about time management." Such promises were usually fulfilled, sometimes with comments added, e.g., "I like Lakein's distinction between pacing and planning because I hadn't really thought of it that way before."

Framework promises, e.g., "There are three major aspects of time management." These promises were usually fulfilled in the form of a series of capsule summaries within the structural framework (cf. essay #005).

Common sense concept, e.g., "Time management is important for everyone." These concepts were sometimes developed, sometimes ignored in favor of more local concepts (cf. essay #100), and sometimes developed for a while and then abandoned for other local concepts or summary.

Moral at the end. Usually a belated promise that the time management advice does indeed add up to something that everyone can learn from (cf. essay #066), this strategy was difficult to evaluate. Sometimes the moral looked like an insight arrived at during writing; in other papers it seemed to be a general technique for rounding off what was basically a summary. Its organic relationship to the text varied widely. In some papers it took the form of a comment on the next-to-last paragraph of material, while in others it related to the source text as a whole or the writer's response to the topic, rather than to the essay.

Unique concept. When this promise was offered, the concept was often signalled by a formal or "sermon" voice (e.g., "We must all learn and practice time management") or by a "written-to-the-examiner" voice, as if to demonstrate mastery of the material, e.g., "Although experts on time management agree on certain basic points, they disagree about the need for rest." In the problematic papers, the unique concept frequently deteriorated into a summary of the source text.

Interpretation for an original purpose. When the writer focused on a group of readers, the text often promised to show struggling students that they would be better off using time management techniques. Other writers used their paper to show how expert advice confirmed "what I've always thought and tried to practice" or for some other purpose of interest to them. Such papers usually carried out the announced purpose, except when they drifted into summary.

Disguised summary, e.g., a summary, as in Essay #066, that closely follows the structure of the source text but is presented as a process, with the ideas seemingly addressing the problems that a student would encounter as they would occur. With some notable exceptions, writers usually added little to these descriptions by way of evaluation or personal experience; that is, they typically did not blend source material with original material and they did not create an original process structure with the source material embedded in it.

On the basis of these observations we developed two new categories—the Isolated Main Point/Conclusion and the Frame—that described some of the hybrid essay structures we were seeing. Essays in the category of Isolated Main Point/Conclusion typically began with what looked like a controlling idea and then lapsed into summary, or, like essay #100, began informally and then used the summary as the basis for a main idea that was stated in the final paragraph. Essay #066 (which was also marked as a synthesis) offers a slightly more sophisticated version of this pattern: It begins with two generalizations—the second introduces the summary of Lakein—and concludes.
with a summary statement about learning to manage time that is unrelated to the topic of the first sentence ("Extensive research has produced many books on the subject of Time Management in professional and educational settings."). Essays in this category relied on local coherence strategies; the larger idea, whether at the beginning or end of the essay, seemed tacked on to what was basically a summary or a summary with comments.

Essays in the Frame category, by contrast, began and ended with the same generalization, e.g., that planning one's time and using it effectively are the two major aspects of time management. The discussion usually developed the framing concept, but the concept was subordinated to a summary of the sources. These categories let us accurately distinguish essays that seemed to the judges to have a predominately review and comment structure plus signals for other sorts of coherence. Because these papers give relatively consistent coherence cues, they may reflect a step on the path of a writer's development. Nevertheless, the attempt did not produce a thorough-going transformation of ideas that supported a sense of conceptual coherence.

This conflict between promise and delivery will always make categorizing such essays difficult. Coordinating the signals and the substance may indeed be an important skill that freshmen writers are trying to learn. However, if these papers are at all representative of college-level work elsewhere, one can conclude that our freshmen had learned the value of coherence and the wisdom of offering their readers promises of coherence. Most of the papers written for the Time Management task made some such promises. The problem for teachers is to decide to what extent the promises are performed.

The existence of the textual inconsistencies discussed in this report suggests that composing with a written source text was difficult for our students. Even when they had an idea that could seemingly give them enough momentum to reinterpret the source material for a clear-cut purpose (e.g., helping their fellow students), the sheer presence of the source text seemed to create a barrier between the students and the paper they wanted to write. This impression was strengthened by the interviews and self-analyses, which indicated that many students thought they had written quite different kinds of papers than the readers saw. Report 4 presents the students' side of the story--the ways that they saw the task and the real difficulties that they faced but sometimes failed to meet.
REFERENCES


25

18


27

20


31

24


Appendix I: Essay Categories and Instructions to Judges

Category 0: Summarize

USUALLY,
the summary consists mostly of gist statements, and possibly a stray sentence or two of authorial comment that may reduce the entire text to one or two sentences.

HOWEVER,
it need not include each main point from the text and it need not follow the plan of the text.

IN ITS LOWEST FORM,
the "summary" can be a short group of sentences that relate to the topic but don't say much.

Category 1: Review and Comment

These papers show some (limited) independence of the text.

OFTEN,
they read like a free-writing response statement.
When comments and summary appear throughout the essay, the summarizing purpose controls the paper (comments are subordinated to it).

ALTERNATIVELY,
comments may appear in only the first or last paragraph, or may focus on the writer's response to only one or two ideas in the text.

IN RARE CASES (the most sophisticated form of this category),
the essay has a controlling idea or theme that emerges during the summary, but which is never stated as being relevant for the entire paper.
These papers read like nicely focussed summaries.

Category 2: Isolated Main Point and/or Conclusion

USUALLY,
the opening contains what could be called a controlling concept; essay then reviews, summarizes, or reviews/comments on the sources without reference to the original main point; the ending may have a conclusion or a moral.

ALTERNATIVELY,
the opening doesn't say much;
the essay reviews, summarizes, or reviews/comments on the sources;
the ending has an original conclusion or opinion.

IN ANY CASE,
the body does not contain links that tie the review/comment to the concept in the beginning/end.
Category 3: Setting up a Frame

USUALLY,
the opening sets up a frame, then uses the structure of the text for summary and/or comment.
The frame can be very open and simple, e.g., "time management has 3 important aspects—planning, environment, and stamina." E.g., "time management produces better quality work." E.g., "Some of this advice could help students."

ALTERNATIVELY,
the framing concept may emerge during the discussion and be stated in the final paragraph.
The frame can be stated in the first person, e.g., "I don't entirely agree with this advice."

TYPICALLY,
the frame creates an expectation of a focused discussion that may occur in a loose (review and comment) manner.

HOWEVER,
The focus may get lost or become implicit by the end of the paper.
As a controlling concept, the frame is so vague that it can hardly be called a unique concept.

Category 4: Free Responses to the Topic

This is a paper written with little or no reference to the time management text.

Category 5: Controlling Concept

ESSAYS IN THIS CATEGORY contain a stated organizing concept which controls the discussion.
Even if the "discussion" is just a glorified summary,
it is subordinated to the concept.

HOWEVER,
the text may thematize authors rather than concepts.
(Watch out for this misleading cue.)
The organizing concept appears throughout the essay and should be clearly identifiable, even when diffused.

RHETORICALLY,
the implied purpose is, usually, to discuss and demonstrate understanding of the text material.

HOWEVER,
the beginning may set up an ostensible rhetorical purpose that is not carried out in the body of the essay.
Category 6: Interpretations for a Purpose

The discussion is subordinated to the writer's purpose, which may be to help the writer herself. The "discussion" may be a rhetorically presented summary or a summary with comments.

IN ALL CASES, the purpose, whether stated or implied, is clearly identifiable and is carried out in the body of the essay.

Decision Path for Judging:

1. Is the essay a summary, with little or no other comment?
   If YES, choose Category 0.
   If NO, choose Categories 1, 2, 3, 5, or 6.

2. Does the discussion contain something that looks like a controlling concept or frame?
   If NO, choose Category 1 (sometimes 2).
   If MAYBE or YES, choose Categories 2, 3, 5, or 6.

3. Does the concept or frame actually control the discussion?
   If NO, choose Category 2.
   If YES, choose Categories 3, 5, or 6.

4. Is the concept a FRAME? That is, does it set up an open structure for discussion that lets the writer a) discuss aspects of the topic and b) use a review structure for the rest of the essay?
   If YES, choose Category 3.

5. Is the concept a "real" CONCEPT? That is, does it commit the writer to a focussed discussion and control the focus of the paragraphs that follow?
   If YES, choose Categories 5 or 6.
   If the concept is a "real" concept but does not control the focus of the paragraphs that follow, choose Category 1.

6. Does the essay reinterpret the assignment for an original purpose and write to a particular audience?
   If YES, choose Category 6.
   If NO, choose Category 5.
Appendix II: Essays Using the Interpret for a Purpose Organizing Plan

This organizing plan took two somewhat different forms which are best illustrated with examples from the freshman texts. In essays such as 115, the apparent organizing purpose is to explore a question, think through a conflict, or to solve a problem. The papers present an issue-based, intellectual purpose stimulated by a conflict, quandary, or a need. In other essays the writer's purpose was expressed as an attempt to address or instruct the reader in some particular way. These papers seem motivated more by an audience-based, pragmatic purpose reflecting an awareness of the special needs of the reader. The essay 045 below is an interesting hybrid in that it is clearly addressed to a particular reader, but also tries to involve that reader by raising a question and potential conflict.

Essay 115
One of the most important and least discussed issues facing students is time management. Based on a number of research and self-help books, I feel that the solution to the problem of time management is a compromise of many of the opinions of professionals whose arguments are similar in some ways, but still disagree overall. Many of their "helpful hints" are actually obstacles to many students whose work habits are similar to mine.

Of those recommendations from research conducted on time management, I feel that the most hindering of these concerns students scheduling large blocks of time in advance in order to organize the work day. Alan Lakein, author of How To Get Control of Your Time and Your Life, is in agreement with Walter Pauk of Cornell University. Both men advocate the use of scheduling work throughout the day in order to subdivide what might appear as a large task. In theory I agree with them, however when a student attempts to accomplish this scheduling, he finds that plans do not usually go as expected. There are various types of interruptions and complications that occur in everyday life; obviously there are too many to be listed here. In case there are no interruptions, many students, who are like myself, will create their own interruptions in the form of procrastination.

There is a paradox resulting from Pauk's suggestion in that when put in a "distraction-free" room, a student will seek out a distraction, either consciously or unconsciously, and thus not concentrate fully on his task . . . [approximately a page of the essay reviewing and evaluating the source ideas follows]

My argument with James' theory is based on what economists call the Law of Diminishing Returns, . . . For example, if a French student were studying the new vocabulary for a quiz, he might learn thirty new words in the first hour, twenty in the second, and ten in the third. He keeps learning, but less efficiently as he continues . . .

After studying some of the research on time management, I have formed my own formula for time management. First of all I agree with the majority of the students who want to "get the easy stuff out of the way," but once that is accomplished, there should be some time left over to work on the long term projects. It comforts me to think that if I work on the project in advance, I can relax near the time it is due; in this case I agree with the "create a crisis
theory."

When I am ready to begin an assignment, I am alert and have the radio on. If, at some point in the assignment, I feel fatigued, or that further study will not result in more knowledge, I take a rest break. By following these few simple guidelines, I and others with similar study habits should be able to effectively manage our time.

Essay 045

Have you ever felt you had too much to do in too little time? Have you ever wondered how you are going to get all your work done? Have you ever stayed up all night long to finish a project or cram for a test? As college students, we've all felt bowled over by our workloads at one time or another. It's unavoidable. There simply are not enough hours in a day. Or is that really the case?

It is no lie that the college level workload puts extreme pressures on a student's time. To escape the time burden, college students try to do the maximum amount of work in a minimum amount of time. We've all got all sorts of tricks to decrease our work time. We do the "easy" stuff first and rush through projects at the last minute. We try to complete writing assignments in one sitting. We gloss over things we don't have to turn in, if we bother to do them at all. We use these and other strategies constantly. But that's how you play the game. It's the only way to do all that is expected of us. Right?

Let's take a good look at ourselves. Sure, we're getting the work done, but it is really as good as it could be? We often sacrifice quality for quantity. Our cute time-saving strategies may decrease the pressure of large assignments, but our overall performance is not up to par. We've got a problem. What can we do? As college students, we need to learn better strategies that not only permit us to use time wisely but to also do our best work.

This key is time-management. [The rest of the § discusses planning time.]

Okay, so now we've found enough study time and scheduled it into our days. That's not enough. We need to use the new found time effectively. Here's where concentration comes into play. A good student working environment is quiet and free of distractions. This may mean turning off the stereo or unplugging the phone. Groan . . . moan . . . but aren't your grades worth it? Another key to concentration is avoiding fatigue. . . .

Time management, when practiced properly, can be a tremendous asset to college students. By learning to use time wisely, we can get maximum performance in minimum time. And that's what we're all looking for.
Appendix III: Inter-rater Agreement on Elaborated Taxonomy

<table>
<thead>
<tr>
<th>Rater Pairs</th>
<th>Kappa</th>
<th>SG</th>
</tr>
</thead>
<tbody>
<tr>
<td>LN/KB</td>
<td>.10</td>
<td>.11</td>
</tr>
<tr>
<td>LN/XR</td>
<td>.39</td>
<td>.13*</td>
</tr>
<tr>
<td>LN/RJ</td>
<td>.25</td>
<td>.09*</td>
</tr>
<tr>
<td>KB/XR</td>
<td>.46</td>
<td>.15*</td>
</tr>
<tr>
<td>KB/RJ</td>
<td>.27</td>
<td>.11**</td>
</tr>
<tr>
<td>XR/RJ</td>
<td>.65</td>
<td>.12*</td>
</tr>
<tr>
<td>All raters</td>
<td>.37</td>
<td>.05*</td>
</tr>
</tbody>
</table>

Percent of Essays Agreed on in Each Set (range): low = .78, high = 1.00; mean = .90
Percent of Unanimous Agreement in Each Set (range): low = .26, high = .56; mean = .40
*p < .01; **p < .05.

A Because we originally used three raters, the essays were divided into six sets, so that each possible pairing of raters would occur. After the revised taxonomy was developed, a fourth rater was added who read all of the essays. We thus ended up with six possible pairings of raters. Although agreement on the revised taxonomy was taken to mean that two out of any group of three raters agreed, in fact the raters were often unanimous.

Discussion of Cohen's Kappa

Cohen's Kappa (Cohen, 1960) is a statistical procedure developed for measuring agreement among judges when the data involves independent subjects, independent ratings by judges, and a nominal scale having categories that are independent, exhaustive, and mutually exclusive. Unlike the better-known r values, the Kappa does not simply measure a relative proportion of agreement/disagreement. The number of an r value represents a proportion of agreement or disagreement that is greater than chance (0.00), such that when the r value is subtracted from ±1.00 (representing absolute agreement or disagreement), the remainder is understood to mean a proportion of noise (i.e., chance, bad data, etc.). Thus, a high r number indicates a large proportion of agreement/disagreement and a small proportion of noise. Cohen's Kappa is a one-tailed procedure that involves calculating the frequency of chance-expected agreement and subtracting it from the observed agreement. A Kappa states the proportion of non-chance agreement.

Unlike an r value, however, a Kappa does not stand on its own as a meaningful measure of agreement because the proportion of chance-expected agreement will vary depending on the distribution (variance) of the data. If the raters' agreements are spread out over all of the categories in the nominal scale, the chance-expected frequency of agreement will be small and the Kappa will be large. If, however, the data is skewed so that the raters' agreements involve few categories, the chance-expected number will be large and the resulting Kappa will be small. In both cases, the actual number of agreements might be the same. As an extreme example, this researcher encountered an instance in which judges agreed on 48 out of 48 items, an agreement that an r value would express as 1.00. Yet because 47 of the 48 agreements were located in a single category, the chance-expected frequency of agreement was so high that the Kappa was .00.

Therefore, the relative size of the Kappa cannot be taken by itself as a measure of agreement. Instead, the value of the Kappa must be considered in the context of data.
distribution. This context is provided by the Standard Gamma, which gives the size of the standard deviation (the variance of the data). A small Standard Gamma indicates narrow standard deviations; a large Standard Gamma indicates wide standard deviations. The ideal is thus to have a relatively large Kappa and a relatively small Standard Gamma, as is the case with the overall interrater reliability shown above. A Kappa that is ±1.96 of its Standard Gamma is significant at the .05 level; a Kappa that is ±2.58 of its Standard Gamma is significant at the .01 level, and so on. Thus a Kappa of .50 might reflect a very high non-chance frequency of interrater agreement, no agreement, or any level of agreement in between, depending on the size of the Standard Gamma. In contrast to $r$ values, for which significance probabilities are not normally given, the Kappa requires a Standard Gamma and a significance probability to be correctly interpreted.