The Writer's Knowledge and the Writing Process: A Protocol Analysis

Peter Smagorinsky, University of Oklahoma

Abstract. This study used on-line protocol analysis to contrast the effects on the writing process of knowledge taught in three instructional treatments: *Models* (declarative knowledge of form), *General Procedures* (declarative knowledge of form plus general procedural knowledge related to content plus procedural knowledge related to form), and *Task-Specific Procedures* (declarative knowledge of form plus task-specific procedural knowledge related to content plus procedural knowledge related to form). Pretest and posttest protocols from six students in each treatment measured treatment effects on the processes of students writing essays involving extended definition. Students in the *Models* treatment made weak improvements in relating the elements of definition and did not think critically about the concepts being defined. Students in the *General Procedures* treatment made gains in linking ideas according to particular task constraints and improved their critical thinking skills. Students in the *Task-Specific Procedures* integrated their ideas purposefully, thought critically about the concepts being defined, and appeared to establish a conversational voice to anticipate composing needs.

Composition authorities have begun to debate the sources of knowledge that most benefit writers. The most ancient notion of all—and still practiced by about a third of the secondary teachers studied by Applebee (1981)—holds that studying models of successful products enables writers to produce the features of the exemplars. Most current theorists agree that instruction in essay form is insufficient and, according to some, counterproductive. Critics have presented two process-centered alternatives to the study of model essays: learning general composing procedures, usually in the form of free-thinking activities such as brainstorming and freewriting; and learning task-specific composing procedures, which vary according to the specific knowledge required to undertake particular tasks.

Hillocks (1984, 1986a, 1986b) has been the foremost advocate of task-specific knowledge, using his meta-analysis of experimental research as the basis for his judgments. He questions the effectiveness of the "natural process" method of teaching writing in which a teacher facilitates writing

I would like to thank Susan Goldin-Meadow, Fred Lighthall, William Pattison, Michael W. Smith, Tom Trabasso, and especially George Hillocks, Jr., plus the retiring editors and anonymous reviewers from *RTE* for their help in preparing this manuscript.

by having students choose their own topics and explore them through unrestricted writing. Donald Murray (1980) describes such instruction: "I would not write—would not need to write—if I knew what I was going to say before I said it. . . . In the writing process approach, the teacher and student face the task of making meaning together. The task is ever new, for they share the blank page and an ignorance of purpose and outcome" (p. 13). Writing assignments by teachers "guarantee bad writing," he says (p. 14). Rather, he maintains,

> we teach our students to write by allowing them to experience the process of writing. That is a process of discovery, of using written language to find out what we have to say. We believe this process can be adapted by our students to whatever writing tasks face them—the memo, the poem, the textbook, the speech, the consumer complaint, the job application, the story, the essay, the personal letter, the movie script, the accident report, the novel, the scientific paper. There is no way we can tell what our students will need to write in their lives beyond the classroom, but we can give our students a successful experience in the writing process. We can let them discover how writing finds its own meaning. (p. 20)

Hillocks, on the other hand, would contend that each of the tasks enumerated by Murray requires specific procedural knowledge. He describes, for instance, Troyka's (1973) experimental treatment for teaching argumentation in which students engage in role-playing activities to learn and practice the task-specific skills of formulating and supporting generalizations, predicting, evaluating, and responding to counter-arguments. In contrast, Hillocks describes a possible way to teach students to write a fable (1986a, pp. 90–91):

> Before asking children to write a fable [a teacher] would have asked them to read several fables over a period of time and then asked them to generalize about the form, particularly the relationships among the key elements. . . . She might have asked students to revise or complete inadequate fables . . . asking them to think of morals, elaborations of how a prideful mouse might act, what such a mouse might say in a given situation, and so forth. She might then have asked children to think of human characteristics that annoy them [and] to think of animals that might be used to represent each such characteristic: a hypocritical cat [and so on]. The teacher might then have asked youngsters to work in groups listing foolish things an egotistical monkey . . . might do, what might happen as a result, what the animal might learn, and so forth. In short, before making the writing assignment, the teacher would help students develop the substantive, formal, and procedural knowledge necessary to writing a fable. Finally, she would have provided opportunities for students to share their ideas and writing before producing a finished draft.

Both Murray and Hillocks agree that knowledge of essay form is inadequate in learning to write, Murray regarding it as harmful and Hillocks seeing it as insufficient. Both scholars have developed alternatives based

on attention to process. Murray advocates engagement in an all-purpose writing process that promotes both the discovery of ideas and the development of good writing. Hillocks sees a place for non-linear thinking and/or writing as part of a process (note the brainstorming session for foolish acts by an egotistical monkey) but structures the activity so that it directs students towards task-specific knowledge such as procedures for addressing counter-arguments in argumentation or procedures for creating personification in writing fables. Murray describes the teacher's role as supportive of students' engagement in writing, but essentially non-directive; Hillocks believes that teachers are most effective when they use their own awareness of task-related knowledge to design and sequence activities for students to engage in, with explicit attention to specific procedures and their purposes. Murray sees the writer discovering purpose in the process of composing; Hillocks sees the writer working towards a goal through the strategic employment of procedures learned through direct instruction.

This intra-discipline dispute over the relative effectiveness of different types of knowledge in improving writing is part of a broader debate in cognitive theory. Scholars have contested the relative effectiveness of general procedures and task-specific procedures in general learning for decades. Some researchers have investigated the general applicability of certain heuristics such as the "General Problem Solver" (Ernst & Newell, 1969; Newell & Simon, 1972) and found them effective procedures for solving new problems under different circumstances. Others have questioned the transferability of general procedures. Glaser (1984; see also Rabinowitz & Glaser, 1985) studied the problem-solving processes of experts and found that they had a knowledge of domain-specific patterns and an ability to recognize and apply their knowledge of these patterns to solve novel problems. Pressley, Snyder and Cariglia-Bull (1987) found that effective thinking depends on specific, context-bound skills and knowledge that have little application to other domains; they find little empirical support for the idea that we can teach transferable, generalizable, context-independent skills and strategies. Some researchers have tried to reconcile the opposing theories by suggesting that a balance of general and task-specific procedures is most beneficial (Perkins & Salomon, 1989) or by investigating conditions when general procedures are amenable to transfer (Nickerson, Perkins & Smith, 1985).

The use of models, too, has had its defenders, even though they have been disdained by recent composition experts (Emig, 1971; Collins & Gentner, 1980; and many others). Research in observational learning (studying models in such varied areas as dance, psychotherapy sessions and problem solving; see Bandura, 1977; Strupp & Bloxom, 1973) has shown that the observation of certain acts may result in better performance than simply attempting to perform such acts. Little Leaguers learning to bunt could thus benefit more from watching an expert model

bunting techniques than if they were simply to engage in "discovery bunting."

We have, therefore, a problem not confined to the field of composition but representing a dispute over the types of knowledge that affect learning in general. This investigation uses protocol analysis to contrast the effects of instruction in models, general procedures, and task-specific procedures on the composing process of students writing essays involving the definition of abstract concepts such as "friendship" or "leadership." Definition serves as an appropriate task for contrasting the effects of different types of knowledge for several reasons. First of all, definition is a skill required by members of diverse disciplines: scientists who create taxonomies, critics who establish criteria for quality, lawmakers who differentiate between acceptable and unacceptable behavior, policy-makers who establish priorities, and so on; and a skill that drives more humble pursuits such as establishing criteria for determining which breakfast cereal to buy or for ranking the nation's top twenty basketball teams.

Furthermore, definition is a task with distinct features which have been accepted for millennia, dating back at least to Aristotle's definition of moral virtues in *Nichomachean Ethics.* As described by Hillocks, Kahn, and Johannessen (1983, p. 276), those who endeavor definition "work from a large body of experience and/or systematically collected data. The phenomena under consideration are first circumscribed generally and then increasingly specified and differentiated through the use of criteria and examples." Instructional books that have taught definition (i.e., D'Angelo, 1977; Johannessen, Kahn, & Walter, 1982) have identified elements that are common to effective definitions: *criteria* that state rules that a candidate for inclusion must meet; *examples* that illustrate the criteria; and *contrasting examples* that seem to illustrate the criteria but lack some essential characteristic. For instance, one student from the present study offered the following sequence as part of a definition of friendship:

> Criterion: I think true friendship might also include taking certain risks or tasks for another person.
> Example: Someone who saves your life is probably going to be your friend for life.
> Elaborate example: That's how the Chinese see it.
> Elaborate example: If you save their life, they owe you a favor for life.
> Elaborate example: A true friend might save your life,
> Contrasting example: and a friend might hesitate because he's more concerned about his own well being than yours.

This student has gone through the same process one might experience in formal tasks such as establishing laws or informal tasks such as forecasting the winners of Oscar awards. The elements of a definition can be illustrated in a model, as I have just provided; or students can learn procedures for producing the elements, both task-specific procedures for

writing definitions and general procedures that could apply to other writing tasks.

The aggregate of experimental research on the types of knowledge contrasted here has suggested that task-specific knowledge taught through the study of discrete data sets (Hillocks' "Inquiry" focus) is more effective in improving writing than general procedures (such as freewriting) or the study of models (Hillocks, 1984, 1986b). What causes these differences? How do students *think* as a result of different instructional foci? The purpose of this study is to investigate writers in the process of composing to examine *why* different types of instruction have different effects on product.

Method

Participants

The protocol study drew a sample of students from a larger study designed and directed by George Hillocks, Jr., and supported by grants from the Benton Center for Curriculum and Instruction at The University of Chicago. The Hillocks study (in progress) will examine treatment effects on the composing products (i.e., their finished essays) of all students in the study; the present study examines treatment effects on the composing process of representative students. The participants came from three team-taught regular track eleventh grade American Studies (English and History) classes at a large suburban high school, with 50–55 students in each class. Participants in the protocol study came from a pool of students determined to be of "average" and approximately equal ability by both their pretests and their regular teachers' assessments. From this pool nine girls and nine boys volunteered to contribute protocols. Fourteen of the students were white and four were African-American, approximately reflecting the school's racial balance.

Treatment Assignment

Each American Studies class was divided into three randomly stratified sub-classes of 16–18 students, with the sorting based on the results of pretest essays in which students defined "courageous action." The scores were determined by students' ability to generate criteria, examples, and contrasting examples. Each sub-class was randomly assigned to one of three treatments—*Models, General Procedures,* and *Task-Specific Procedures*—each treatment occupying 12 days of instructional time. Three expert teachers were randomly assigned to the three treatments for the first class and assigned to the remaining two classes so that across all three, each teacher taught each treatment once. The treatment groups within each class met separately so that each was in effect a separate class. This coun-

terbalancing of teachers and treatments controlled for teacher and treatment order effects.

To review, then, the participants were evenly distributed by gender, treatment, and teacher. From each of the three American Studies class sub-classes, protocols were collected from one boy and one girl from the *Models* treatment, one boy and one girl from the *General Procedures* treatment, and one boy and one girl from the *Task-Specific Procedures* treatment.

Prior Instruction

The study took place in October–November of a typical school year. In the two months of school prior to the study, the teachers of the students who contributed protocols taught writing in the following ways: Two of the teachers taught a general expository model based on the five paragraph theme; rather than using essay models, they provided an outline of the theme's form, then required students to submit finished essays within about three days. The third teacher gave weekly writing assignments usually involving analysis of literature, with the content based on class discussions. None of the three teachers reported instruction in procedures; they all felt that elaborate instruction in writing would take time away from covering the course content.

Treatments

All three treatments included study in model essays with about 40 percent of the instruction uniform in all treatments. The purpose of this common instruction was to examine the effects of manipulations in different treatments following instruction in a common source of knowledge. Following the identical introductory sequence, the treatments varied as follows (see Appendix A for complete descriptions of the treatments):

Models

Students in the *Models* treatment studied, labeled, and evaluated additional models of definition essays. Their instruction did not present explicit procedures for generating the elements of a definition (criteria, examples, and contrasting examples). Rather, they studied exemplars such as the following excerpt from a model illustrating a criterion and appropriate support for "terrorism" (Hillocks, in progress):

> *Criterion statement:* The targets of terrorism are government and civilian populations. Whereas acts of war may result in the accidental killing of civilians, terrorism involves harming civilians on purpose.
> *Example that fulfills the criterion:* The 1983 bombing of the U.S. Embassy in Beirut, Lebanon, is an example of a terrorist act. The targets were ambassadors, government officials, and civilians working in the

> embassy. For the attackers, the embassy was a symbol of the enemy—the United States.
> *Contrasting example—one that is close to the term defined but does not fulfill the criterion:* On the other hand, if civilians are killed by guerrilla fighters in the process of bombing enemy soldiers launching offensive operations, the action cannot be labeled terrorism. In this case, the killing of civilians is accidental; the target is the enemy soldiers.

They then evaluated a series of definitions ranging from the simple ("A truck is a device on wheels used for moving loads") to more complex concepts such as "cruelty to animals." Finally they wrote essays defining "courageous action."

General Procedures

Students in the *General Procedures* treatment were taught the procedures of freewriting and brainstorming to generate criteria, examples and contrasting examples, plus strategies for revision. For instance, they were given illustrations of how to brainstorm and freewrite to define such concepts as "democracy," and then used the procedures to define concepts of their choice and ultimately "courageous action."

Task-Specific Procedures

Students in the *Task-Specific Procedures* treatment were taught the procedure of studying problematic examples to generate criteria, examples and contrasting examples, plus strategies for revision. For instance, to define "freedom of speech," students studied a set of seven problematic examples and were instructed to generate a criterion from each one (some criteria perhaps governing more than one example). The examples presented controversial issues such as: "Orson Welles produced a radio dramatization of H. G. Wells' *War of the Worlds,* written as news flashes announcing an invasion of the United States from Mars. The program was so realistic that it caused panic on the East Coast. Orson Welles was fired." Students defined additional concepts through the study of problematic examples, and ultimately wrote essays defining "courageous action."

Evaluation Task

For the pretest, half of the subjects wrote on *leadership* and half wrote on *friendship,* topics which a pilot study had found to be comparable in difficulty to *courageous action* (which served as both a topic for a separate pretest and a definition topic in all three treatments). The pretest topics were reversed for the posttest. Protocols were collected from both pretest and posttest essays using standard protocol procedures (Smagorinsky, 1989a; Swarts, Flower & Hayes, 1984) with the analysis measuring change following instruction. The students produced their protocols during the school's 42-minute class periods concurrent with the production of essays by their

classmates for the Hillocks (in progress) study. Most students finished within the 42-minute period.

Protocol Analysis

Segments

I first divided the protocols into "segments." A segment is a section of text that states a single idea (such as a single criterion or example). Most often, a segment consists of a single sentence.

Coding

After dividing the protocols into segments, I coded each one for the specific operations the writers used when producing an essay involving definition. Thus, the system categorizes the *content* found in the definition essays, including different types of *generalizations* such as criteria; *support,* such as examples and contrasting examples; and the ways that writers tied generalizations and support together *(warrants).* I also identified *processes* such as when writers conducted memory searches, referred to the essay's form, referred to their conception of the task, and so on. A trained rater segmented and coded one protocol from each treatment for both pretest and posttest with the following rates of agreement: segments .92, categories .86. The complete coding system is described in Table 1.

I have previously illustrated how writers use *criteria* in extended definitions. At this point I should distinguish between *criteria* and what I call *attempted criteria* since important treatment effects depend on this distinction. As described earlier, a *criterion* distinguishes between members and nonmembers of the concept being defined, with members illustrated by an example and nonmembers illustrated by a contrasting example. Here, for instance, is a criterion developed by Jim (General Procedures Treatment) in his posttest protocol: "Friends tend to spend time together; weekly, monthly, etc. Some friendships are just associates, such as classmates which out of class you do not spend time with." In his criterion Jim clearly distinguishes between members (people with whom one spends time by choice outside the workplace) and nonmembers (those with whom one does not spend time by choice). Often students would attempt to generate criteria without making such distinctions. Had Jim merely said, "Friends tend to spend time together," he would not have made such a clear distinction; the time spent together could have been obligatory. Generalizations such as this I labelled *attributes.*

A second type of attempted criteria I called an *incidental statement;* that is, a generalization which, while true, is not definitional. On his pretest, for example, Jim had written, "Most, instead of leading, choose to follow and those who can lead others become leaders." This statement is proba-

Table 1

Coding Categories

The protocols were coded into the following categories. The rate of agreement follows each category name.

Content

Criterion

(.75): General rule essential to making a clear distinction between members and nonmembers. A criterion either states a contrast, or is supported by a contrasting example that discriminates between one who meets the definition and one who only seems to. Thus, the contrast must be between closely related behaviors, rather than opposites. A criterion is determined by the way the student uses a generalization to make distinctions, not by how well a generalization matches the researcher's notion of the concept. Example:

> True friends do things together after school, whereas acquaintances might be friends in class but never do anything together outside school.

Attribute

(.97): Characteristic behavior of the person measured against the concept; it is often a potential criterion. It lacks the contrast necessary to distinguish members and nonmembers. Example:

> Friends do things together.

Incidental statement

(.65): Statement that, while usually true, is nonessential, evasive, tangential, preconditional, consequential, related to the topic but not definitional, or synonymous but not definitional. Example:

> Without leaders, the world would be in chaos.

Perspective

(1.): Statement asserting that creating standards is relative. Example:

> You can't say what's right and wrong for everyone.

General Contrast

(.6): Contrast that is not specific enough to include or exclude members from the concept. Example:

> Many relations are based on what is thought to be friendship but actually is not.

(Continued)

Table 1 (continued)

Example

(1.): Particular behavior to illustrate an attribute or criterion. Example:

> Leadership can be in business, like you have a meeting and you're trying to get something accomplished.

Contrasting Example

(1.): Example of behavior that lacks some essential characteristic and therefore does not satisfy a criterion. Example:

> If a team did not listen to a coach's advice and trust what he said, but just did what they wanted, the team would not succeed.

Warrant

(.5): Explanation of how an example illustrates a criterion or attribute. Example:

> For example, if a religious leader persuades others to donate money, for his own personal use and then uses it for his own personal needs, [warrant begins] than he is doing it for the wrong reasons, and he is taking advantage of those people who trusted his leadership.

Summation

(.67): Statement at the end of the essay that brings closure to the essay, summarizing or synthesizing the ideas from the paper. Example:

> Using these characteristics you may be able to distinguish the difference between leadership and true leadership.

Process

Judgment

(1.): An assessment of ideas produced; can be positive or negative. Example:

> No, that's stupid.

False start

(1.): Line of thought cut off early, and then abandoned. Example:

> True leadership would be someone who'd be. . . .

Block

(.9): Writer is unable to generate ideas. Example:

> Let me see, I'm stuck right here. I can't get into this.

(Continued)

Table 1 (continued)

Search

(1.): Statement that initiates a review of the writer's knowledge. A writer may search for content, examples or criteria. Example:

> I'm going to think of times in my own life when I knew people weren't my friends and tried to pretend they were.

Form

(.83): Statement about the structure of the essay, either the essay as a whole *(macro)* or at the word/sentence level *(micro)*. Example:

> macro: I think I'm going to start my main, my body. I'm thinking that I need to use what I said in my opening paragraph, refer to that.
> micro: I think I spelled that wrong.

Task Conception

(1.): Statement referring to any of the writer's understanding of the task constraints and requirements, including the assignment, the instruction, the rater's expectations, or staying on topic. Example:

> This paper doesn't look like it would be asking for a personal experience.

Determination

(.8): Statement concerning a decision or observation regarding the content of the essay. This can include decisions to reread, revise, or recopy the essay; decisions to include an idea in the essay or to leave an idea in the essay intact. Example:

> I should probably proofread it.

Peripheral statement

(1.): Statement peripheral to thinking about or generating the essay, including questions to the researcher on how to perform the protocol, statements about time limitations, off-topic remarks about the research project, judgments about the difficulty of the protocol process itself, or statements about the essay's neatness. Example:

> Are we the only English class doing this?

bly true (if circular) but does not help to define the concept. Yet it is an attempt at generating a criterion.

Attempted criteria emerged as important in the protocol analysis for two reasons. First of all, they gave an indication of students' understanding of the relationship among definition elements; students who generated attributes and incidental statements tended to use them *as though they were* criteria, supporting them with examples or contrasting examples.

Second of all, by studying the ratio of criteria to attempted criteria I could examine the extent to which students improved their ability to make distinctions and thus think critically about the definition task.

Analyses

Following segmenting and coding, I analyzed the data by isolating each category and tracking it through each protocol to identify patterns in students' thinking, looking in particular for patterns of interaction between and among categories. (See Smagorinsky, 1989b, for the full data analysis.) The data from the protocols were extraordinarily rich, allowing for a study of a great range of composing processes. Due to the limited number of participants, however, many of the trends, while interesting, were not strong enough to be conclusive. The results reported here focus on areas in which treatment effects were clearest. In each case, I will present my qualitative analysis of the protocols, followed by ANOVAs and *t* tests to examine the statistical significance of between-group differences.

Results

General Treatment Effects and Hypotheses

The protocols revealed differences among the treatment effects in two main areas which I call *purposeful composing* and *critical thinking.*

Purposeful Composing

By *purposeful composing* I mean the extent to which students relate generalizations (including criteria, attributes, and incidental statements) to support (including examples and contrasting examples). Students related generalizations to support in one of four ways:

1. Supporting a criterion with both an example and a contrasting example.
2. Supporting a criterion with only an example.
3. Supporting a criterion with only a contrasting example.
4. Supporting an attempted criterion (an attribute or an incidental statement) with an example.

I made the following hypotheses regarding students' purposeful composing:

1. Students in the Models treatment would improve their ability to relate generalizations to support because their instruction illustrated such relationships in the exemplars.
2. Students in the General Procedures treatment would improve their purposeful composing through the study of models; their instruc-

tion in freewriting and brainstorming, however, would not substantially improve this skill further because the procedure does not necessarily engage them in establishing an explicit relationship between generalization and support.

3. Students in the Task-Specific Procedures treatment would improve their purposeful composing through the study of models; the task-specific procedure of generating criteria from problematic examples would further improve this skill because the procedure required students to develop a relationship between the two elements and generate distinctions between closely-related behaviors.

Critical Thinking

By *critical thinking* I mean: (1) students' ability to improve their ratio of criteria to attempted criteria; and (2) students' ability to reject attempted criteria instead of including them in their essays. I made the following hypotheses regarding students' critical thinking:

1. Students in the Models treatment would show little improvement because their instruction did not include procedures for discriminating between criteria and attempted criteria; rather, they simply learned where to place a generalization in the essay structure.
2. Students in the General Procedures treatment would improve their critical thinking because they learned procedures for generating criteria; the free-association strategies, however, would not consistently lead them to discriminate between strong and weak generalizations.
3. Students in the Task-Specific Procedures treatment would improve their critical thinking more than students in the General Procedures treatment because they were taught a procedure that required differentiation among closely related behaviors in order to generate strong criteria.

Illustration of Treatment Effects

The opening portions of the pretest and posttest protocols of Sonya, a student from the Task-Specific Procedures treatment, illustrate how a writer could improve according to measurements of *purposeful composing* and *critical thinking*. Each segment is initially identified by its coding category; italicized segments are those written in the essay.

Pretest

False Start: Friendship is—
Search: What can I write?
Determination: I can write about my friends.
Determination: OK, I know what to do.
Incidental Statement: *Friendship is a powerful thing.*

Negative Judgment: That sounds stupid, it is stupid.
Negative Judgment: No, I don't like that.
Block: I don't know what to write about.
Task Conception: Friendship . . .
Determination: I'll just write that.
False Start: Friendship is—
Search: How is it powerful?
Block: Well, because . . . I don't know.
Incidental Statement: Let's see. Friendship is powerful because it could hurt you—OK, *Friendship is a powerful thing because it could hurt you.*
Negative Judgment: Wait, I don't like that.
Perspective: *Friendship means many things to many people.*
Positive Judgment: Yeah, that sounds better.
Reread: Friendship means many things to many people.
Search: What else should I write?
False Start: OK, friendship is—
Negative Judgment: Whoops, that doesn't make sense.
False Start: Friendship is—no, wait.
Perspective: People think of friendship in many ways. Hum. *People think of friendship in many ways.*

Sonya goes on in this uncertain manner to produce a relatively long protocol (145 segments) in which she produces one criterion, six attributes and fourteen incidental statements, all of which she writes in the essay; for these generalizations she produces four related examples, one related contrasting example and two unrelated contrasting examples.

Posttest. Her posttest reveals great changes in her composing process:

Task Conception: Leadership—think of an analytical definition first.
False Start: Leadership is following someone—
Negative Judgment: No.
Attribute: Leadership is leading a group and—leading a group—no, yeah, *Leadership is leading a group and helping them through*—and helping them through *situations.*
Search: OK, now criteria.
Search: No, first I should think of an example.
Example: *A preacher is a leader of a church.*

Elaborate Example: *He or she guides people through the session and—*

Negative Judgment: No, that's stupid.

Determination: I'm going to tear that sheet up [starts a new sheet of paper].

Attribute: *Leadership is someone in a group who is more realistic—*

Attribute: Who is more intelligent?

Negative Judgment: No,

Elaborate Attribute: is more realistic, yeah,

Attribute: *and more confident.*

Negative Judgment: No, I don't like that either [starts a new sheet of paper].

Criterion: *Leadership is a person in a group who*—a person in a group who, who everyone extends—*everyone listens and pays attention to more.*

Example: *For example*—OK, for example *a leader in a gang would be considered leadership.*

Elaborate Example: *He tells the people of his group what to do* and then—wait—he tells the people of his group what to do and they react to his—*and they do what he tells them.*

Warrant: *They listen to he or she.*

Search: OK, now I have to think of a contrasting example.

Reread: They listen to him and they do what he tells them.

Contrasting Example: *Leadership would not be if some people picked a person who was not*—wasn't *confident in themselves and the person*—and the person *told the people he or she do what they please.*

Warrant: *That would not*—wouldn't *be considered leadership, because*—hum—because *people would not pay attention to a person like that in the first place because the person has to be confident*—well *at least sound confident and not be scared.*

Determination: OK, that's one.

False Start: Leadership is—

Negative Judgment: No, that's not in a group.

Incidental statement: Ok, leadership is also—*Leadership is also participating in some kind of social, political or*—social, political—no, what's the word—yeah, social, political or *economic event.*

Example: *For example,* people—oh, I know—*high school students go on a field trip to*—I don't know—to *the Smokey Mountains and they have certain rules.*

Criterion: *People who stick to those rules have some leadership.*

Elaborate Criterion: They stick together—*They have to stick together and follow the rules.*

Elaborate Criterion: *Leadership* would not—*wouldn't be considered if people didn't want rules and just be free.*

Contrasting Example: *For example, anarchists, people who don't want a government.*

Sonya's posttest protocol consists of 61 segments, in which she rereads her essay and produces an additional example and contrasting example for her second criterion. Her protocol reveals the generation of two criteria (both supported by examples and contrasting examples), three rejected attributes, and six written incidental statements.

Effects

Purposeful composing. The change between her pretest and posttest protocols indicates great improvement in both purposeful composing and critical thinking. Her composing becomes more purposeful because, instead of casting about aimlessly for things to say as she does in the pretest, she uses her knowledge of form and procedures to produce criteria and support in relation to one another, often using a memory search to produce appropriate content. Her composing has become methodical, perhaps a perjorative term to some who favor unrestricted writing, but an approach that serves her well under these task constraints and at this point in her development as a writer. We would hope that with more practice these processes would become more automatic, thus enabling her to attend to other aspects of thinking and composing.

Critical thinking. Note that a single segment can illustrate two treatment effects. The criteria she generates in the posttest illustrate *purposeful composing* when viewed in relation to the supporting examples and contrasting examples; they also illustrate *critical thinking* when viewed in relation to the number of attributes and incidental statements she produces. Sonya improves her critical thinking by improving the ratio of criteria to attempted criteria from 1/20 (both total instances and those written in the essay) to 2/9 (total instances) and 2/6 (written in the essay); and she improves the percentage of attempted criteria written in the essay from 100 percent (20/20) to 67 percent (6/9).

Additional effects. The change in Sonya's writing process also reveals other treatment effects (although not conclusively different from one treatment

to another) which I would not have predicted, yet which made a great deal of sense following the protocol analysis. I had predicted that students would increase the length of their protocols from pretest to posttest due to their greater knowledge of the task requirements. Most students, however, regardless of treatment, *reduced* the length of their protocols (four of six students in each treatment). Those students who reduced the length of their protocols tended to do so substantially: Of students reducing the length of their protocols, students in the *Models* treatment averaged a 36 percent reduction, students in the *General Procedures* treatment averaged a 40 percent reduction, and students in the *Task-Specific Procedures* treatment averaged a 56 percent reduction.

Similarly, I expected students to increase the number of criteria generated due to the instructional emphasis on generating criteria, but many students in fact generated fewer criteria in the posttest than in the pretest. A close reading of the protocols indicated that in the pretests, students floundered in their attempts to think of criteria and examples as did Sonya. In the posttests students had a much better grasp of the task and tended to go about it more systematically. Thus, instead of thinking of a great many criteria, they would generate fewer criteria and spend more time thinking about how to support them with appropriate examples; and instead of producing longer protocols, they often took a more purposeful approach to the task and accomplished it more efficiently. The clear differences among treatments, then, came in the *relationship* of elements and in the *processes* (such as memory searches) that appeared to improve students' ability to relate them, rather than in the discrete accumulation of definition elements.

Statistical Evidence

Purposeful Composing

To measure the change in purposeful composing I calculated the ratio of generalizations (criteria, attributes, and incidental statements) which were unsupported by examples or contrasting examples to generalizations which were supported. I will illustrate how I determined an *improvement score* with the performance of Tony from the Task-Specific Procedures treatment. On his pretest protocol Tony had produced a total of 11 unsupported generalizations and zero supported generalizations for a ratio of $^{11}/_{0}$; on his posttest protocol he produced five unsupported generalizations and seven supported generalizations for a ratio of $^{5}/_{7}$. I defined "improvement" as decreasing unsupported generalizations and improving supported generalizations. Tony's improvement score, then, was calculated by anticipating a *decline* in the numerator and therefore assigning a positive value to the amount of decline ($11-5=6$) and anticipating an *increase* in the denominator and therefore assigning a positive value to the

amount of increase (7 − 0 = 7). Tony's improvement score was the sum of these two values (6 + 7 = 13).

Table 2 reports contrasts of improvement scores for purposeful composing for the three treatments, with the results of a one-way ANOVA with three levels. Scores labelled *Written* represent the portion of the protocol which the students entered in the final draft of their essays; scores labelled *Total* represent the total number of segments in the protocol, both written and nonwritten.

According to one-tailed *t* tests for individual comparisons among treatments, students in the Task-Specific Procedures treatment scored significantly higher than students in the Models treatment for both the Total ($t = 2.659$; $p < .05$) and Written ($t = 1.855$; $p < .05$) measurements of purposeful composing.

Critical Thinking

To measure the change in critical thinking I calculated the ratio of attempted criteria to criteria. I will use Tony again to illustrate how I determined an *improvement score.* On his pretest protocol Tony had generated seven attempted criteria and two criteria for a ratio of ½; on his posttest protocol he generated three attempted criteria and two criteria for a ratio of ⅔. I defined "improvement" as decreasing attempted criteria and increasing criteria. Tony's improvement score, then, was calculated by anticipating a *decline* in the numerator and therefore assigning a positive value

Table 2

ANOVA For Improvement Scores: Purposeful Composing

Treatment	Means	*SD*
Total		
Models	2.167	6.853
General Procedures	8.333	10.132
Task-Specific Procedures	10.166	2.714
Written		
Models	3.5	7.064
General Procedures	5.5	9.628
Task-Specific Procedures	9.333	3.077

Source of Variation	*DF*	Mean Square	*F*	Significance of *F*
Total				
Group	2	105.389	2.014	.168
Error	15	52.333		
Written				
Group	2	52.722	1.04	.378
Error	15	50.689		

to the amount of decline ($7-3=4$) and anticipating an *increase* in the denominator and therefore assigning a positive value to the amount of increase ($2-2=0$). Tony's improvement score was the sum of these two values ($4+0=4$).

Table 3 reports contrasts of improvement scores for critical thinking for the three treatments, with the results of a one-way ANOVA with three levels.

According to one-tailed t tests for individual comparisons among treatments, students in the Task-Specific Procedures treatment scored significantly higher than students in the Models treatment for both the Total ($t=2.588$; $p<.025$) and Written measurements ($t=2.584$; $p<.025$); and students in the General Procedures treatment scored significantly higher than students in the Models treatment for both the Total ($t=2.405$; $p<.025$) and Written ($t=2.086$; $p<.05$) measurements of critical thinking.

Percentage of Attributes and Incidental Statements Generated but not Written

I would expect effective instruction to give students procedures for identifying and editing weak generalizations. Evidence of such discrimination appeared when students thought of a weak generalization but did not write it in their essays. In the pretests, students were likely to think of an attribute or incidental statement and write it in their essays; following instruction they tended more to reject these weak generalizations. Students

Table 3

ANOVA For Improvement Scores: Critical Thinking

Treatment	Means	*SD*
Total		
Models	−.167	5.811
General Procedures	7.667	5.465
Task-Specific Procedures	8.0	5.099
Written		
Models	0	5.762
General Procedures	5.667	3.327
Task-Specific Procedures	7.167	4.0

Source of Variation	*DF*	Mean Square	*F*	Significance of *F*
Total				
Group	2	128.167	4.29	.034
Error	15	29.878		
Written				
Group	2	85.722	4.49	.03
Error	15	19.078		

from the Models treatment wrote down 91 percent of their attributes and incidental statements on the pretest and 81 percent on the posttest; students from the General Procedures treatment wrote down 61 percent on both pretest and posttest; and students from the Task-Specific Procedures treatment wrote down 79 percent on the pretest and 59 percent on the posttest. Students in the Models treatment entered the study with a poor ability to discriminate and gave evidence on the posttest of a still-poor but improved ability; students in the General Procedures treatment entered with a relatively good ability and did not improve it; students from the Task-Specific Procedures treatment entered with a poor ability and gave evidence of improvement towards a relatively good ability.

Discussion

This study, within its limitations, has suggested how knowledge taught in different instructional treatments affects the ways in which writers think as they compose according to task-determined constraints. For centuries teachers have assumed that if they illustrate essay structure students will be able to express their own ideas in that form independently. Eschholz (1980) has written that "Certainly few people will take exception to the general rule that one good way to learn how to write is to follow the example of those who write well" (p. 21). He goes on to say that English teachers who instruct through the imitation of models

> feel secure talking about the important themes contained in the various reading selections; discussing diction, figurative language, sentence structure, and paragraph patterns; classifying prose readings into the traditional categories of description, narration, exposition, and argumentation; and correcting student essays for syntax, spelling, punctuation, and style. Such activities go well with lecture-discussion courses which meet as a class three to five times a week; with the various textbooks . . . that are available; and with the skills of the majority of English teachers who have been trained to teach literature and perhaps grammar, but unfortunately not composition (p. 23).

Eschholz's characterization of the models method reveals much about its appeal and problems. Its popularity stems from its expediency and compatibility with teacher-centered instructional approaches. Yet these very factors that make it attractive account for the problems it creates. The role of the teacher is to explain what the finished product should look like; the role of the students is to comprehend the nature of the assigned subject matter, understand how to render that comprehension into the type of writing illustrated by the model, figure out how to structure their response to a topic so that it fits into the form dictated by the model, and finally transform their ideas into writing. For the teacher this method is indeed expeditious since the analysis of the models takes little class time; yet for the students the task is onerous since they must figure out

how to generate appropriate ideas and represent them in the assigned form.

This study suggests that students who are taught solely through the study of model essays have a great deal of trouble teaching themselves how to write in the fashion of their instructional exemplars. Students in this study who were taught only with models neither learned the structure of the essay particularly well nor thought clearly about the ideas they generated. If we can picture a model essay as a form in which a writer slots information, students who studied models did not effectively learn the placement of the slots and did not think critically about the ideas that they put in the slots.

The study of models combined with instruction in composing procedures appeared more efficacious for these student writers. Both the general procedures of freewriting and brainstorming and the task-specific procedures of generating criteria from problematic examples appeared to give students strategies for producing the particular elements stressed in the model essays, with task-specific procedures having a stronger effect on improving students' purposeful composing and with both procedural treatments enabling students to think critically about the concepts being defined.

These findings present some difficult problems for practitioners. As noted earlier, two of the three of the students' regular teachers—both well-respected in their school—taught through general composing models because such instruction occupies relatively little class time. Alternatives that involve procedural knowledge, while promoting more effective composing strategies, take up a great deal of time. The general procedures approach—such as that described by Murray (1980)—requires that teachers devote considerable time to allowing students to engage in the writing process. Teachers are under pressure from different sources to achieve a variety of goals: cover course content, teach writing, teach vocabulary, teach grammar/syntax, teach reading comprehension and literary understanding, promote cultural literacy, encourage global consciousness, prepare students to take standardized tests and so on. Expanding time devoted to writing necessarily takes time away from these other obligations.

Hillocks' theory presents an even greater problem because of the instructional time needed for each type of writing. The sequences for argumentation, definition, and other types of writing are different and might take several weeks each; most teachers are reluctant to sacrifice so much class time for attention to the composing process when they can cover more material by assigning literature, discussing it, and then giving a writing assignment, perhaps in imitation of some writing model.

Beyond that, instruction in task-specific procedures requires teachers to spend a great deal of their planning time thinking about the knowledge one needs to engage in particular writing tasks, identifying procedures that writers can learn to produce them effectively, and designing activities

that enable students to learn and practice the procedures. In spite of the great intellectual challenge of this task, we can imagine that the conception and design of, for instance, several series of problematic examples might take more time than many teachers are willing to spend. Widespread instruction in task-specific procedures would require either an extraordinary commitment by practitioners to the development of materials, or the training of a knowledgeable generation of teacher-consumers who would support a market for the commercial production of texts based on understanding of task-specific composing procedures.

Instruction in general composing procedures might be a good alternative in that they are relatively effective in improving writing yet not quite so demanding on teachers' time. For some tasks, such as personal expression in which writers discover their own topic, task, and procedures, general procedures are perhaps the ideal mode of instruction. For more constrained tasks such as definition, students probably need more than simple engagement in the act of writing. Students in this study combined the use of general procedures with the study of model essays, an interaction which clearly boosted the power of the models instruction and, if we accept the results of Hillocks' meta-analysis of experimental research (1984, 1986b)—in which the models focus emerged as stronger than the freewriting focus—probably strengthened the effects of the general procedures. For tasks with a clear structure, models appear to help illustrate relationships among ideas and increase the effectiveness of the procedures.

Instruction in task-specific procedures, in spite of the seemingly prohibitive amount of time one must spend developing and teaching them, appear to have an important place in the English class as well. Hillocks' findings on the effectiveness of task-specific procedures—the "inquiry" focus in which teachers identify tasks (i.e., definition, argument) that students might fruitfully engage in, and teach task-specific procedures that will enable students to write effectively about self-chosen topics—are well-documented (1984, 1986b). Students in this study who combined the study of models with instruction in task-specific procedures improved in thinking more purposefully and critically about their task and topic, perhaps due to the development of a conversational voice to direct their thinking to appropriate content knowledge. When Rod, for instance, defined friendship on his posttest he generated the criterion that "True friendship requires sacrifices. When worst comes to worst, you are there. This does not mean the situation has to be bad in order to make a sacrifice." His next segment is an explicit self-cue to initiate a memory search for relevant information: "OK, now I definitely need an example." He then sorts through several possible examples from his experience before generating a hypothetical situation in which someone's friend gets AIDS and is banned from school; the person sticks up for his friend even though his loyalty will probably cost him a student council election. Stu-

dents who were taught task-specific procedures showed a greater tendency to engage in this type of thinking than did students in the other treatments, even though the differences were not statistically significant. Task-specific procedures, then, appear to promote composing which is driven by purposeful, critical, and strategic thinking, qualities which are certainly beneficial for students.

This study has helped look into composing processes of students exposed to instruction emerging from three competing, although not incompatible, theories. The clearest result is that the study of model essays alone does little to help students improve their process of essay production. While economical for time-penurious teachers, the study of models puts the great burden of learning how to write on students, who are apparently not up to the task. Teachers would benefit from providing some sort of procedural instruction to give students a method for transforming their content knowledge into coherent written expression.

References

Applebee, A. (1981). *Writing in the secondary schools.* National Council of Teachers of English (Research Rep. No. 21). Urbana, IL: National Council of Teachers of English.

Bandura, A. (1977). *Social learning theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bereiter, C., & Scardamalia, M. (1982). From conversation to composition: The role of instruction in a developmental process. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2, pp. 1–64). Hillsdale, NJ: Erlbaum.

Collins, A., & Gentner, E. (1980). A framework for a cognitive theory of writing. In L. Gregg & E. Steinberg (Eds.), *Cognitive processes in writing* (pp. 51–72). Hillsdale, NJ: Erlbaum.

D'Angelo, F. (1977). *Process and thought in composition.* Cambridge, MA: Winthrop.

Emig, J. (1971). *The composing processes of twelfth graders* (National Council of Teachers of English (Research Rep. No. 13). Urbana, IL: National Council of Teachers of English.

Ernst, G. W., & Newell, A. (1969). *GPS: A case study in generality and problem solving.* NY: Academic Press.

Eschholz, P. A. (1980). The prose models approach: Using products in the process. In T. R. Donovan & B. W. McClelland (Eds.), *Eight approaches to teaching composition* (pp. 21–36). Urbana, IL: National Council of Teachers of English.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist, 39,* 93–104.

Hillocks, G. (1984). What works in teaching composition: A meta-analysis of experimental treatments. *American Journal of Education, 93*(1), 133–170.

Hillocks, G. (1986a). The writer's knowledge: Theory, research, and implications for practice. In A. R. Petrosky & D. Bartholamae (Eds.), *The teaching of writing* (pp. 71–94). Chicago: The University of Chicago Press: The National Society for the Study of Education.

Hillocks, G. (1986b). *Research on written composition.* Urbana, IL: National Council of Teachers of English.

Hillocks, G. (in progress). Acquiring knowledge for writing: The processes and effects of three focuses of instruction.

Hillocks, G., Kahn, E., & Johannessen, L. (1983). Teaching definition strategies as a mode in inquiry: Some effects on student writing. *Research in the Teaching of English, 17,* 275–284.

Johannessen, L., Kahn, E., & Walter, C. (1982). *Designing and sequencing prewriting activities.* Urbana, IL: National Council of Teachers of English. (ERIC Document Reproduction Service No. ED 211 982)

Murray, D. (1980). Writing as process: How writing finds its own meaning. In T. R. Donovan & B. W. McClelland (Eds.), *Eight approaches to teaching writing* (pp. 3–20). Urbana, IL: National Council of Teachers of English.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nickerson, R., Perkins, D. N., & Smith, E. (1985). *The teaching of thinking.* Hillsdale, NJ: Lawrence Erlbaum.

Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher, 18*(1), 16–25.

Pressley, M., Snyder, B. L., & Cariglia-Bull, T. (1987). How can good strategy use be taught to children? Evaluation of six alternative approaches. In S. M. Cormier & J. D. Hagman (Eds.), *Transfer of learning* (pp. 81–120). New York: Academic.

Rabinowitz, M., & Glaser, R. (1985). Cognitive structure and process in highly competent performance. In F. D. Horowitz & M. O'Brien (Eds.), *The gifted and talented: Developmental perspectives* (pp. 75–98). Washington, DC: American Psychological Association.

Salomon, G., & Globerson, T. (1987, October). Rocky roads to transfer. *The second annual report to the Spencer foundation.* Tel-Aviv, Israel: Tel-Aviv University.

Smagorinsky, P. (1989a). The reliability and validity of protocol analysis. *Written Communication, 6*(4), 463–479.

Smagorinsky, P. (1989b). *The effects of different types of knowledge on the writing process: A protocol analysis.* Unpublished doctoral dissertation, The University of Chicago.

Strupp, H. H., & Bloxom, A. L. (1973). Preparing lower-class patients for group psychotherapy: Development and evaluation of a role induction film. *Journal of Consulting and Clinical Psychology, 41,* 373–384.

Swarts, H., Flower, L., & Hayes, J. R. (1984). Designing protocol studies of the writing process: An introduction. In R. Beach & L. Bridwell (Eds.), *New directions in composition research* (pp. 53–71). NY: Guilford Press.

Troyka, L. Q. (1973). *A study of the effect of simulation-gaming on expository prose competence of college remedial English composition students.* Unpublished doctoral dissertation, New York University.

Appendix A: Treatments

All three treatments included the following:

1. A general introduction to definition, including reasons for generating precise definitions.
2. The purpose and audience for definitions, including several perspectives on definitions for life and death as exhibited in court cases such as the Karen Quinlan case.
3. The parts of extended definitions, including analytic definitions, criteria, examples, and contrasting examples, with warrants included in the models although not so labeled. Students identified the parts in model essays on terrorism and prejudice.

4. An explanation and illustration of the four parts of an analytic definition: the word defined, a linking verb, classification, and differentiating characteristics. Analytic definitions were modeled, and then students evaluated given definitions to judge their quality. They then wrote analytic definitions of such words as "touchdown" and "religion," following this with peer feedback.

The treatments varied as follows:

Models

Students in the *Models* treatment were taught declarative knowledge of form, identifying criteria, examples, and contrasting examples in model essays and evaluating essays to determine whether someone else had written a definition well. They were not taught procedures for generating original criteria or support. Their particular lessons were:

1. In small groups, students judged the quality of essays defining profound curiosity, cruelty to animals, and loyalty.
2. Students evaluated court cases to determine how well lawyers defined and supported their clients' cases.

The next five activities were included in both the Models and the General Procedures treatments:

3. Students evaluated model essays to determine how well a phenomenon (hallucination) was defined and illustrated.
4. Given an essay about sharks and one about whales, students determined whether each was a definition, labeling any definition parts found in each essay.
5. Students examined given definitions (loyalty, news, etc.) and evaluated them as too broad or too narrow.
6. Students labeled the parts of a model essay on courage.
7. Students defined rules in given sports (i.e., what is a fair ball in baseball?).

General Procedures

Students in the *General Procedures* treatment were taught declarative knowledge of form (the study of model essays), procedural knowledge related to substance (the general procedures of brainstorming and freewriting to summon and transform content knowledge), and general process knowledge (knowledge of and practice in revision strategies).

The lessons in this treatment were:

1. Students were taught general "prewriting" strategies for generating ideas about definitions, including:
   A. Brainstorming, which was defined and illustrated in a model of a student brainstorming about defining democracy. Students then brainstormed for a topic to define and for ideas related to defining it.
   B. Freewriting, which was defined and illustrated in a model of a student's freewriting about defining democracy. Students then freewrote for a topic to define and for ideas related to defining it.
2. Students received instruction in revision strategies and peer feedback.

The next five activities were included in both the Models and the General Procedures treatments:

3. Students evaluated model essays to determine how well a phenomenon (hallucination) was defined and illustrated.
4. Given an essay about sharks and one about whales, students determined whether each was a definition, labeling any definition parts found in each essay.
5. Students examined given definitions (loyalty, news, etc.) and evaluated them as too broad to too narrow.
6. Students read model essays on courage, labeling the definition parts.
7. Students defined rules in given sports (i.e., what is a fair ball in baseball?).

The next two activities were included in both the Models and Task-Specific Procedures treatments:

8. Students ranked three models (on profound curiosity, cruelty to animals and loyalty) from best to worst, identified problems with each and then chose one for revision.
9. Students evaluated model essays to determine how well a phenomenon (hallucination) was defined and illustrated.

Task-Specific Procedures

Students in the *Task-Specific Procedures* treatment were taught declarative knowledge of form (the study of model essays), general process knowledge (knowledge of and practice in the general process of revision) and procedural knowledge related to substance (the task-specific procedure of examining examples to generate criteria).

The lessons in this treatment were:

1. Students were taught a particular strategy for defining: evaluating items from a data set (scenarios) according to whether the characters' behavior does or does not illustrate a given concept (freedom of speech in the first activity), and generating a criterion from each scenario. They then did a similar exercise to generate a definition for courage, producing an essay defining courageous action.
2. Students were given a model of a person's thought process while generating criteria, examples, and contrasting examples to define "true curiosity." They then defined generosity in small groups.
3. Students evaluated nine examples of individual action to determine whether they met the criteria for a "Giraffe Award" given to people who stick their necks out to benefit others.
4. Students defined terms (loyalty, news, etc.) and differentiated them from similar terms (i.e., loyalty from blind obedience, news from gossip).
5. Given a series of episodes from which to choose (such as picking the "best" student representative to meet with a visiting king), students defined the concept in a small group and then compared definitions with other groups.

The next two activities were included in both the General Procedures and the Task-Specific Procedures treatments:

6. Students received instruction in revision strategies and peer feedback.
7. Students ranked three models (on profound curiosity, cruelty to animals and loyalty) from best to worst, identified problems with each and then chose one for revision.