Is writing as difficult as it seems?

C. MICHAEL LEVY
University of Florida, Gainesville, Florida

and

SARAH RANSDELL
Florida Atlantic University, Ft. Lauderdale, Florida

Experiment 1 assessed the time and effort allocated to writing subprocesses while generating written and verbal protocols over 10 weekly writing sessions. Within a 40-min session, planning time consumed about 45% in the first 5 min, but stabilized at near 30% thereafter. Generating text initially consumed 40% of the writers' time, peaked at 50% midway, and then declined to its original level. The time spent revising and reviewing was negligible early in writing sessions, but increased substantially late in the sessions. The highest and lowest quality documents could be differentiated on the basis of the amount of time the writers devoted to revising and to the magnitude of their RTs in a secondary interference task. Writers showed consistent, distinctive patterns of transitional probabilities between writing subprocesses both within and across sessions, yielding quantitative representations of their writing styles. In Experiment 2, writers overestimated the amount of time they devote to revising and overestimated the amount of effort they allocate to planning and text generation. Their estimations did not improve after 10 weeks of composing. A time-and-effort-based analysis of writing is proposed to account for these data.

Both professional and nonprofessional writers often lament that the process of writing is arduous and complex. Perhaps not coincidentally, writing is also one of the least understood language-production tasks. One step toward better understanding of any complex production task lies in a detailed and fine-grained analysis of its components. This is one of the main goals of our research.

Three basic writing subprocesses have received extensive theoretical and empirical attention: planning, translating, and reviewing (Hayes & Flower, 1980). With some notable exceptions (Kellogg, 1987, 1988; Kellogg & Mueller, 1993), little empirical work has been done to determine how time and effort are devoted to each of these subprocesses during writing. It would be extremely useful in guiding the development of models that can predict and explain quality writing to have a clearer account of how people allocate their resources to the processes that constitute the task. We need to identify the possible effects of generating verbal protocols (Ericsson & Simon, 1980, 1984, 1993; Stratman & Hamp-Lyons, 1994), currently

our main source of data about writing. And finally, despite the fact that writing is an ongoing activity for many adults, no data have been reported about how, or whether, an individual writer's pattern of writing changes across writing sessions.

In the following sections, we describe research on writing processes from the time-course-and-effort studies that Kellogg and his colleagues began and compare it with our own approach. We then explain the need for longitudinal studies of writing that can inform us about the stability and distribution of quality writing within and across multiple writing sessions.

The Time Course and Effort of Written-Language Production

Theories of written-language production that capture the second-by-second processes of the task are in their infancy. Limited time-and-effort data necessary for developing detailed theories have been collected, but only using indirect means within a single writing session. For example, Kellogg (1987, 1988; Kellogg & Mueller, 1993) periodically probed writers to retrospect about whether they were engaged in generating text, reviewing, or planning. A general pattern of time allocation emerged across these studies. Subjects devoted about 50% of writing time to generating text throughout, but as the session progressed, they spent increasingly less time planning, and more time reviewing. These data were derived at least partly from subjects' assessments of their own processing, which may not always be veridical (Ericsson & Simon, 1980). The present study compares this general picture of time allocation with direct on-line measures of typing behavior supplemented by verbal-protocol information. Because verbal

reports are not evaluative, they are less susceptible to writers' own theories of the process (Ericsson & Simon, 1993).

While they composed, Kellogg's subjects performed a secondary task, responding to a randomly presented tone. If subjects have substantial cognitive capacity reserves while performing the primary writing task, they should be able to allocate an appropriate amount to performing a secondary task, enabling them to perform it faster or better (e.g., Tyler, Hertel, McCallum, & Ellis, 1979). Kellogg (1987, 1988; Kellogg & Mueller, 1993) found that the least effort was required during text generation, that an intermediate level was required during reviewing, and that the greatest effort was required during planning. Kellogg (1987) suggests that text generation is most like the relatively automatic process of speech and should, therefore, require the least effort of all writing subprocesses. Effort was inferred from the interference response times (IRTs) to the tones (i.e., the algebraic difference between the RT to a tone presented while writing and the RT collected during a baseline period). We implicitly assume that the RT task and the various writing subprocesses all draw on common resources, while recognizing that this assumption is still the focus of some controversy (see Fodor, 1983).

The data in the present study bear on two competing hypotheses regarding effort and writing performance. Kellogg's (1987) *workload hypothesis* suggests that better writers should exert less effort during writing than lesser writers, as defined in terms of amount of topic knowledge. For example, a better writer may be able to allocate less effort to text generation, liberating cognitive resources for revising, and thereby improve writing quality. Britton and Tessor (1982) alternately suggest that experts should exert greater effort than novices and hence have less flexibility in allocating resources among the writing subprocesses. The latter idea is closer to Bereiter and Scardamalia's (1987) view that better writers raise the level of difficulty of the task with increasing experience, changing a simple knowledge-telling approach into a more complex knowledge-transforming one. According to a knowledge-transforming view, better quality documents should be written with greater effort than poorer ones if total writing time is held constant. The methods we report here yield a more fine-grained data set than is possible with directed retrospection, and thus may permit more sensitive comparisons to be made.

In contrast to Kellogg's methodology that requires retrospection, we use a direct measure of time that people devote to various writing subprocesses (Levy & Ransdell, 1994). Writers' word-processed documents and verbal protocols are captured on videotape, played back in real time, and coded for various writing subprocesses. The resulting data enable time-and-effort profiles of documents to be studied on a second-by-second basis. Such data can facilitate the development of a process explanation of the time and effort allocated to writing subprocesses through time.

A Model of Writing Processes Based on Three Types of Protocols

The bulk of research on writing processes comes not from direct behavioral measures (e.g., RT, time on pro-

cess), but from verbal-protocol analysis (Bereiter & Scardamalia, 1987; Bereiter, Burtis, & Scardamalia, 1988; Hayes & Flower, 1980, 1983). The present study integrates written, verbal, and resource-allocation protocols to characterize time and effort on writing processes. Written and verbal protocols facilitate operationally defining four basic writing subprocesses: planning, generating text, reviewing, and revising. Written protocols provide a record of all of the text as it unfolds and is edited, including far more than just the final written product. Verbal protocols synchronized in time with unfolding written text include comments that allow us to determine whether pausing that occurs during writing is associated with reviewing text already generated or with planning future text. We also subdivide reviewing time into reviewing and revising epochs because these two subprocesses can be readily differentiated on the basis of a combination of writers' comments and their actual typing behavior.

The contents of verbal protocols provide a rich description of the subprocesses writers use when they write. For example, the widely cited chapter in Hayes and Flower (1980) describes three main processes that writers talk about when they generate protocols: *planning, generating sentences,* and *reviewing.* Planning includes creating ideas, organizing ideas, and setting goals to achieve during composition. Sentence generation refers to the processes of actual language production—selecting words and constructing sentences. For the present studies, we use the term *text generation* for two reasons: (1) our unit of analysis is the text that evolves each second; and (2) our quality assessments are based on the entire text, rather than on individual sentences. Reviewing includes evaluating text already in place as well as editing errors. Our protocol-analysis technique allows us to distinguish between what are clearly different subprocesses—namely, reviewing and revising.

Hayes and Flower (1983) argue that verbal protocols provide direct information about these subprocesses through a working-memory monitor system. Information maintained within working memory drives the contents of one's written and verbal protocols as well as how one juggles the constraints of writing subprocesses. We therefore collect written protocols (real-time records of word-processed writing) and verbal protocols (what writers say while they are writing) to develop converging evidence for a time-course analysis of writing. We also collect resource-allocation protocols (IRTs linked to specific writing behaviors), which index the cognitive effort allocated. The contribution of time-and-effort information regarding writing subprocesses is necessary for further development of a process model.

The Need for Longitudinal Studies of Writing

It is an empirical question whether the distribution of time and effort devoted to planning, sentence generation, reviewing, and revising seen within a 40-min period is constant over several months or more on the same type of writing task. We simply do not know whether the patterns reported earlier, based typically upon a single writing session, will be stable over time.

Because the initial weeks of the present study include the addition of thinking aloud and responding to a tone as a measure of cognitive effort, the design also permits the influence of these conditions on the "natural" writing task to be established. Ransdell (1995) found that generating concurrent verbal protocols can slow the rate of writing relative to silent controls, but does not affect the nature of the task. Ericsson and Simon (1993) review many research findings showing that problem-solving tasks such as writing are not changed by generating concurrent verbal protocols if they already involve a verbal component.

A related question concerns the predictive validity of writer profiles created from written, verbal, and resource-allocation protocols. Can high- and low-quality documents be distinguished on the basis of time-and-effort allocation to writing subprocesses? Can such documents be distinguished by different patterns of transitions between writing subprocesses? Process profiles that address these questions and that can predict writing quality over time have important theoretical as well as practical significance.

EXPERIMENT 1

The purpose of Experiment 1 is to determine the time course and effort of four writing subprocesses: planning, generating text, reviewing, and revising, both within and across multiple writing sessions. We compare writers' time on subprocesses under several conditions in a longitudinal design: while writing alone, writing while responding to a tone (a measure of cognitive effort), and then by six weeks of writing while both responding to a tone and generating verbal protocols. The main goals of the experiment are: (1) to determine unique profiles of time course and effort allocation for relatively higher and lower quality compositions; (2) to investigate the impact of concurrent verbal protocols on such profiles; and (3) to discover the stability of time course and effort allocation within a writing session and across many sessions.

Method

Subjects. A group of 10 undergraduates from the University of Florida volunteered to participate in a 12-week writing study. They were recruited by posters placed around campus inviting people who enjoyed writing and were proficient in the use of a word processor to consider participating. The subjects indicated that they had used a word processor 3–4 h per week for the last 1.5 years, that they wrote frequently for pleasure as well as to meet academic requirements, and that, except for in-class assignments, they did all of their composing on their word processors. They were not paid for their services, but were told that the writer of the single best document would be awarded a $100 gift certificate from a local merchant. The subjects' average VSAT score was 630.

Apparatus and Materials. The writers performed their assignments on a personal computer. As detailed in Levy and Ransdell (1994), the display screen was divided into panels representing three programs running in a Windows environment. The left three-quarters of the screen contained a writing area running the Windows Notepad, a low-end word processor offering cursor control, searching and replacing, and cutting and pasting of text, but no formatting capabilities or spelling and grammar checking facilities. The lower right panel showed a digital clock, from which we determined when sub-

jects made various oral and typed responses. The upper right panel contained the window for the tone-monitoring task, which was driven by a special-purpose program that (1) randomly presented the default Windows "beep," (2) captured through the serial port the subjects' RT to depress a footswitch, and (3) displayed a running count of the auditory signals delivered (Levy, Fryman, & Ransdell, 1994).

The signal from the computer to the monitor was also sent to a special-purpose device (SimulScan) that converted the signal from a VGA signal to an NTSC signal. This enabled us to videotape the subjects' screens unobtrusively. During sessions requiring a concurrent verbal protocol, the subject's voice was also recorded simultaneously on the videotape.

Procedure. Writing sessions began with a 2-min period during which the subject sat with closed eyes, listening for a beep that the computer generated. The beeps were spaced an average of 30 (±15) sec apart. The subjects were instructed to depress a footswitch as quickly as possible whenever they heard the tone. This initial 2-min period served as a baseline for evaluating RTs collected while the subjects wrote. Because the average of the four RTs collected during each session did not differ systematically across weeks, the average of the 40 RTs collected during the 10 weeks of writing was used as the basis for determining each writer's IRTs.

The writers each wrote on a new title each week and had no advance information about the topic. They were challenged to create a document that would be appropriate for submission to the editor of a sophisticated national magazine such as *The New Yorker*. After collecting the baseline RTs, the experimenter announced the title of that day's assignment and the writer began a 40-min composition period. The titles were open ended and designed to enable writers to approach the topic from different perspectives, using whatever genre they wanted. Examples are "The Greatest High of All," "Perfect Job," and "Politically Correct." Weeks 1 and 12 were devoted to determining the subjects' metacognitions about their own writing (described in Experiment 2).

The first two weeks of writing (Weeks 2 and 3) accustomed the writers to the laboratory environment and to the requirement for composing against a deadline. While the computer generated beeps throughout the writing segment, the subjects were told to ignore them. All subjects wrote on the same topics ("The Lottery" and "Just for the Fun of It").

During the third week of writing (Week 4), the subjects for the first time monitored and responded to the beeps that were presented throughout the writing. At the beginning of the session on Week 5, the subjects were instructed in the concurrent verbal-protocol technique and shown a 2-min video that demonstrated its use. Essentially, they were encouraged to talk aloud while they were writing, verbalizing whatever came into their minds and not introspecting or justifying what or how they were writing. Sessions during next 6 weeks were identical to those of Week 4, except for the additional requirement that the subjects provide the concurrent verbal protocol throughout the writing period. The experimenter prompted the writer with "keep talking" whenever there was a silence of about 10 sec.

Week 11 was equivalent to Week 4, except that the subjects wrote for only 20 min, after which they watched the videotape of their writing session and generated retrospective verbal protocols as their composition unfolded. The analyses of these retrospective protocols are not included in this report.

Coding writing subprocesses. Twelve student assistants were trained in using Eventlog (Henderson, 1989) to score sample tapes until they achieved a satisfactory degree of reliability. Eventlog is a program that enables a computer to emulate an n-channel event recorder. The user initially defines keys to represent behaviors of interest. When the program is launched, a rater depresses and releases appropriate keys to signal the onset and offset of the specified behaviors with millisecond resolution. Afterward, the research can inspect the data on screen as virtual multi-channel ink-tracing or can store the data as an ASCII file for further processing.

At least two assistants used Eventlog to score each of the writers' tapes for Weeks 4–11. Levy and Ransdell (1994) explain how the Eventlog records were combined across judges.

The written-protocol records were derived from judges' coding of the following behaviors exhibited on videotapes of the original composition: instances of typing, revising, pauses in typing or revising of various lengths and at various places in the text, moving the cursor through the text, and paragraphing. To force judgments to be based solely upon the visual record of the writing session, the auditory track was silenced. Reliability of coding the visual written record was high, ranging from $r = .83$ for pausing to $r = .95$ for paragraphing. In all measures of reliability and in the statistics that followed, behaviors and the processes they reflected were determined for each 1-sec period.

The tapes were independently re-scored by other trained assistants, who closely monitored the auditory track on the videotapes, but were told to code the verbalizations in terms of what they could see occurring in the writing. Reliability for the audio record based on the verbal protocols was also high; all rs were .90 or higher for the four main categories of planning, rereading, speaking and writing same content, and writing content.

Once both the visual (written) and auditory (verbal) tracks were coded in Eventlog files, we combined the judgments using the rules displayed in Table 1 to determine the moment at which each writing subprocess began and ended. A computer program then determined what occurred during each of the 2,400 sec in a writing session and which writing subprocess was under way whenever each tone occurred. Interrater reliability was quite high, with two judges agreeing over 90% of the time on planning, text-generating, and revising subprocesses.

Assessing writing quality. The documents produced by each writer were individually printed in a standardized format, combined with those of all other writers, sequenced in random order, and presented to a pair of judges who provided blind ratings. These judges were trained undergraduates, whose background and writing experiences were similar to the writers'. Each judge independently scored each document on 13 dimensions of writing quality using a tool developed in holistic quality assessment as part of university English placement examinations (Ransdell & Levy, in press). These dimensions included content, purpose, style, word choice, organization, and mechanics. The judges were instructed to rate all documents on one dimension before repeating the process with the next dimension. Interrater reliability across all dimensions was $r = .90$.[1] The correlations between dimensions ranged from .20 between mechanics and content to .45 between style and organization, indicating partial independence of the subscales from one another. The quality score assigned to a document was the sum of the ratings of both judges, expressed as a percentage of the maximum possible score.

The judges also blindly rated two essays retyped from *The New Yorker* magazine to serve as a professional-quality reference point. In addition, the judges blindly rated 10 documents produced by relatively novice adult writers in a previous experiment by the authors (Ransdell & Levy, 1994). These documents were written under sim-

ilar conditions as those in this experiment, except for the secondary tasks of responding to a tone and generating concurrent protocols.

Results and Discussion

Writing during Weeks 1–3 was intended solely to accommodate subjects to the tasks, and data from these sessions were not analyzed. We begin by showing how collecting thinking-aloud protocols and IRTs did not interfere with subjects' writing, and then discuss time-and-effort allocation patterns within and across sessions. All results described as reliable meet or exceed the 5% level of confidence.

The influence of thinking aloud. As anticipated, the effects of thinking aloud were negligible. For both planning and text generation, the absolute differences in IRTs on Weeks 4 and 5 were less than 50 msec [all $ts < 1.0$, n.s.]. During the time allocated to text generation, writers created equivalent numbers of words while providing concurrent verbal protocols as they did when writing alone, either across sessions, within a session, or within the first 10 min of a session [all $Fs < 1.0$]. As unnatural as thinking aloud first seemed to our writers—most needed considerable coaxing and reminding to keep talking during the first few minutes of Week 5—they very quickly overcame any inhibitory effects that this additional processing imposed.

Because thinking-aloud protocols do not change how writers generate text, we can use them to disambiguate writing subprocesses. Planning time decreased reliably when the written and verbal protocols were combined [$F(1,59) = 129.01$, $MS_e = 254.8$, $p < .001$]; much of this decrement resulted from the reassignment of pauses in the visual record from the planning to the reviewing subprocess. Revising involved more time after using both the auditory and visual records than when using the visual record alone [$F(1,59) = 17.81$, $MS_e = 314.2$, $p < .001$]. Because the reviewing-process category is meaningful only when the auditory track is considered, it was not possible to measure reviewing before Week 5, when the verbal protocols were introduced.

Cognitive effort of subprocesses. IRT is the algebraic difference between the subject's RT to a tone while engaged in a writing subprocess and the RT during the previous baseline period. Both within and across sessions, in comparison with their planning IRTs, writers exhibited longer IRTs while revising [$t(50) = 2.58$, $p < .02$] and generating text [$t(50) = 1.96$, $p < .05$], although the absolute differences were small—about 91 msec. The IRTs

Table 1
Combinational Response Patterns for Determining Writing Subprocesses

Responses From Writing Protocol*	Responses From Verbal Protocol†	Writing Subprocess
Pausing or starting new paragraph	Not rereading	Planning
Anything	Future planning	Planning
Typing	Anything	Text generating
Deleting, making meaningful or nonmeaningful changes, or any cursor movement	Anything	Revising
Pausing	Rereading or critiquing	Reviewing

*Scored from visual track.   †Scored from audio and visual tracks.

were remarkably stable otherwise within and across writing sessions [all Fs < 1.0].

Time on subprocesses. The IRTs were independent of subprocessing time. None of the correlations relating IRT to the time allocated for a subprocess was significantly different from zero, indicating that these variables were measuring different aspects of the writing process.

Within writing sessions, the allocations of time varied systematically. The subjects spent about 45% of their time planning their documents during the first 5 min of the sessions and gradually stabilized thereafter at around 35% [$F(7,371) = 10.25, MS_e = 704, p < .001$]. As the experiment weeks progressed, they devoted less time to planning [$F(35,371) = 1.65, MS_e = 704, p < .01$]. Between the first and last weeks, their planning times declined from 40% to about 30%.

The time spent generating text increased reliably within sessions [$F(7,371) = 7.65, MS_e = 1,562, p < .001$], but not across sessions. The subjects spent about 40% of their time generating text during the first 5 min of the sessions; this later increased to nearly 50%, and declined in the last 5 min back to 40% at the end, yielding significant quadratic and cubic components [$F(1,53) = 40.36, MS_e = 1,464, p < .001$ and $F(1,53) = 3.82, MS_e = 1,026, p < .05$, respectively].

Reviewing time occurred during only about 2%–3% of the first 5 min, rising to nearly 5% for the duration of the sessions, producing a significant main effect [$F(7,371) = 5.23, MS_e = 409, p < .001$]. Slightly less time overall was devoted to revising, but the pattern of time allocation within sessions was nearly the same for these subprocesses. The main effect of time block for revising was also reliable [$F(7,371) = 16.78, MS_e = 354, p < .001$].

In the writing sessions that occurred during Weeks 4–10, the subjects spent about 40% of their time planning and 45% generating text, but only 8% reviewing, 6% revising, and 1% doing other activities. Kellogg (1987) found a similar time allocation for generating text (50%) but a lower percentage for planning (30%) and a higher percentage for reviewing (35%). About 65% of total composing time was allocated to planning by Gould (1980), who operationalized planning as that part of the writing session not devoted to text generation.

When all subprocesses were included in an analysis of variance (ANOVA) of time allocation, a significant three-way interaction emerged between subprocesses, writing sessions, and time blocks within sessions [$F(105,1113) = 1.89, MS_e = 1000.3, p < .001$]. This complex interaction is portrayed in Figure 1 as a series of scatterplot matrices (SPLOMs) that disregards the differences in magnitude between subprocesses. Superimposed on the SPLOMs are the best-fitting straight lines describing the relationship between time allocations between and within sessions during Weeks 5–10. Upward-pointing lines indicate that more time was spent on the process within a specific time block as the weeks passed. Steeper lines reflect greater relative increases in time allocated. The general picture for planning shows progressive decreases within sessions across

TIME ON WRITING PROCESS
Minutes Since Receipt of Topic
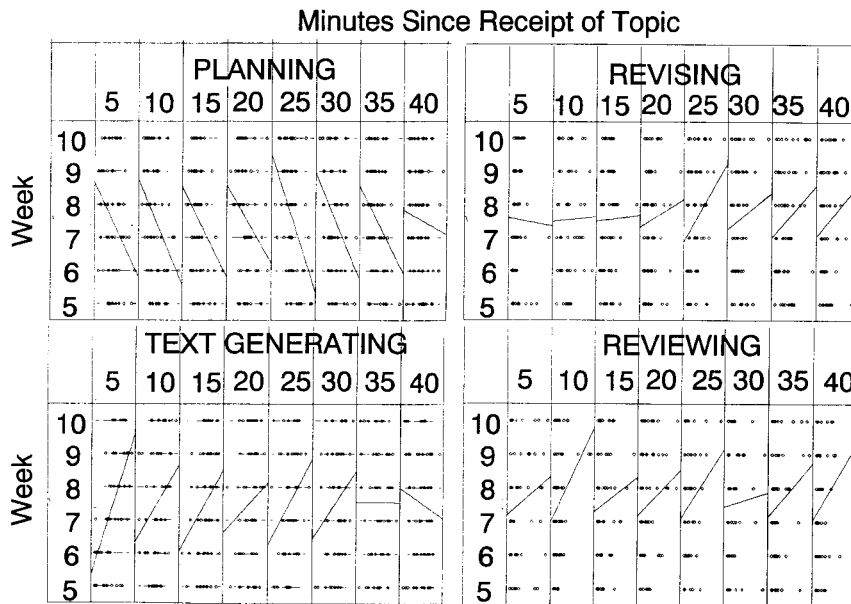
Figure 1. The three-way writing subprocess × week × 5-min time block interaction plotted as a series of scatterplot matrices (SPLOMS) generated by Systat (Wilkinson, Hill, Welna, & Birkenbeuel, 1992). Each row in each cell displays the distribution of times for the 10 writers within each 5-min time block. Each cell also shows the best-fitting straight line for the percentage of time on the subprocess as a function of a particular week and 5-min time block for a specific writing subprocess.

weeks. In contrast, text-generation time increases sharply within the initial 5 min, but increases more slowly across weeks during Min 6–30. Eventually (during Min 31–35), text-generation time became stable, but it began to decline across weeks during the final minutes. Revising time exhibited little change across weeks during the early components of the writing period, but increased across weeks during the later portions of the sessions. Across weeks, reviewing time appeared to increase noticeably at three points: after the writer had composed for about 5 min, and again after 20 and 30 min of writing had occurred.

Writers were remarkably self-consistent from week to week in their differential allocation of time to planning and text-generating across weeks. For all pairs of adjacent weeks, the r values between planning times were about .60–.70, and for text generating, they were about .80. The patterns of revising and reviewing times were less consistent (adjacent-week r values were generally less than .10). However, the multiple R values revealed that between 80%–95% of the variance for each subprocess during Week 10 was accounted for by the times allocated during the prior 5 weeks. Such highly consistent patterns of allocating time led us to look carefully for variables that would help us to better understand individual differences among writers. Some of the results of this search appear in the next two sections.

Processing epochs and transitional probabilities. Epochs were defined by the onset and offset of a writing process. The surprisingly brief median durations of writing subprocesses varied reliably: text-generating epochs were about 7.5 sec, and planning, reviewing, and revising each averaged about 2.5 sec [$F(3,78) = 82.73$, $MS_e = 4.42$, $p < .001$]. Independent of epoch duration, we observed that the number of epochs also varied significantly between the processes. Subjects typically produced about 46 planning and 37 text-generating epochs, but only 15 reviewing and 21 revising epochs [$F(3,78) = 81.17$, $MS_e = 266.8$, $p < .001$].

When epoch duration is ignored, it is possible to analyze the shifts by the writers from one writing process to another by studying tables of transitional probabilities. These probability matrices were determined by dividing the number of occasions on which a writer moved from one subprocess to another subprocess by the total number of epochs within a particular time period. In the present study, we used 10-min blocks of time as the base in order to have 100 or more epochs for any subject. For example, if during the first 10-min block a subject shifted from planning to text generating 55 times and from text generating back to planning 39 times, and within that time block the writer exhibited 105 writing epochs of at least 1-sec duration, the transitional probability for planning→text generating would be .53 (55/105), and for text generating→planning it would be .37 (39/105). Note that these probabilities are independent of the duration and number of epochs. For example, if all of this subject's 11 remaining epochs reflected shifts between revising and reviewing (.10), this subject would have twice the probability of revising→ reviewing than would someone who also shifted 11 times between these two processes but whose session contained

a total of 210 shifts between processes. These probabilities provide a means to capture the recursiveness in a subject's writing.

Figure 2 shows the transitional-probability matrices for two subjects that capture their performance during 10-min time blocks on Weeks 5 and 10. The top left matrix displays the transitional probabilities for the first 10 min in the writing period by Subject 1 on Week 5. The cells are filled with various shades of gray that indicate the magnitude of the probability of shifting from one writing subprocess to another. The diagonal cells are undefined. The top row represents the probability of the subject shifting from planning to text generating, reviewing, and revising. The left column shows the probability of changing subprocess from text generating, reviewing, or revising to planning.

On Week 5, Subject 1 initially shifted between planning and text generating or reviewing about equally often, and between planning and revising less so. Over the 40-min period, there was a gradual and stable redistribution of transitional probabilities, so that the writer shifted from one writing process to another nearly equally often. This is reflected in a tendency for all cells in the matrix to be homogeneously gray. The four matrices for this subject on Week 5 closely resemble their counterparts on Week 10, and are clearly different for those exhibited by Subject 5.

Subject 5 oscillated frequently from planning to text generating and back within each 10-min block on Week 5. Six weeks later, she was still exhibiting a similar set of transitional-probability patterns. This subject almost never shifted from reviewing to text generating or revising or from revising to reviewing. The transitional probabilities for other writers were similar in showing high consistency within days and across weeks. Most writers had distinctive features in their transitional-probability matrices that distinguished them from those produced by other writers. These writing "signatures" (Levy & Ransdell, in press) characterize how subjects micromanage their recursive shifts among writing processes.

The documents produced by Subject 5 on both weeks illustrated were among our lowest in rated quality. The documents produced by Subject 1 were among our highest. While the transitional probability matrices among writers of the highest quality documents differed from one another, they tended to show relatively homogeneous patterns, like those displayed for Subject 1. In contrast, when the majority of the shifts in subprocesses occurred as they did for Subject 5, between planning and text generating and vice versa, the quality of the document suffered.

To our knowledge, transitional-probability matrices have not been used previously to characterize writers' performance. At a minimum, their utility as a research tool depends upon their revealing above-chance levels of performance. If participants randomly shifted from one writing process to another during a specific time block on any day, each of the 12 off-diagonal cells in the transitional probability matrix should be equiprobable. (The diagonal cells are not meaningful here because they reflect the momentary probability of a writer continuing to be engaged with the same process.) In most cases, the sum of the probabil-
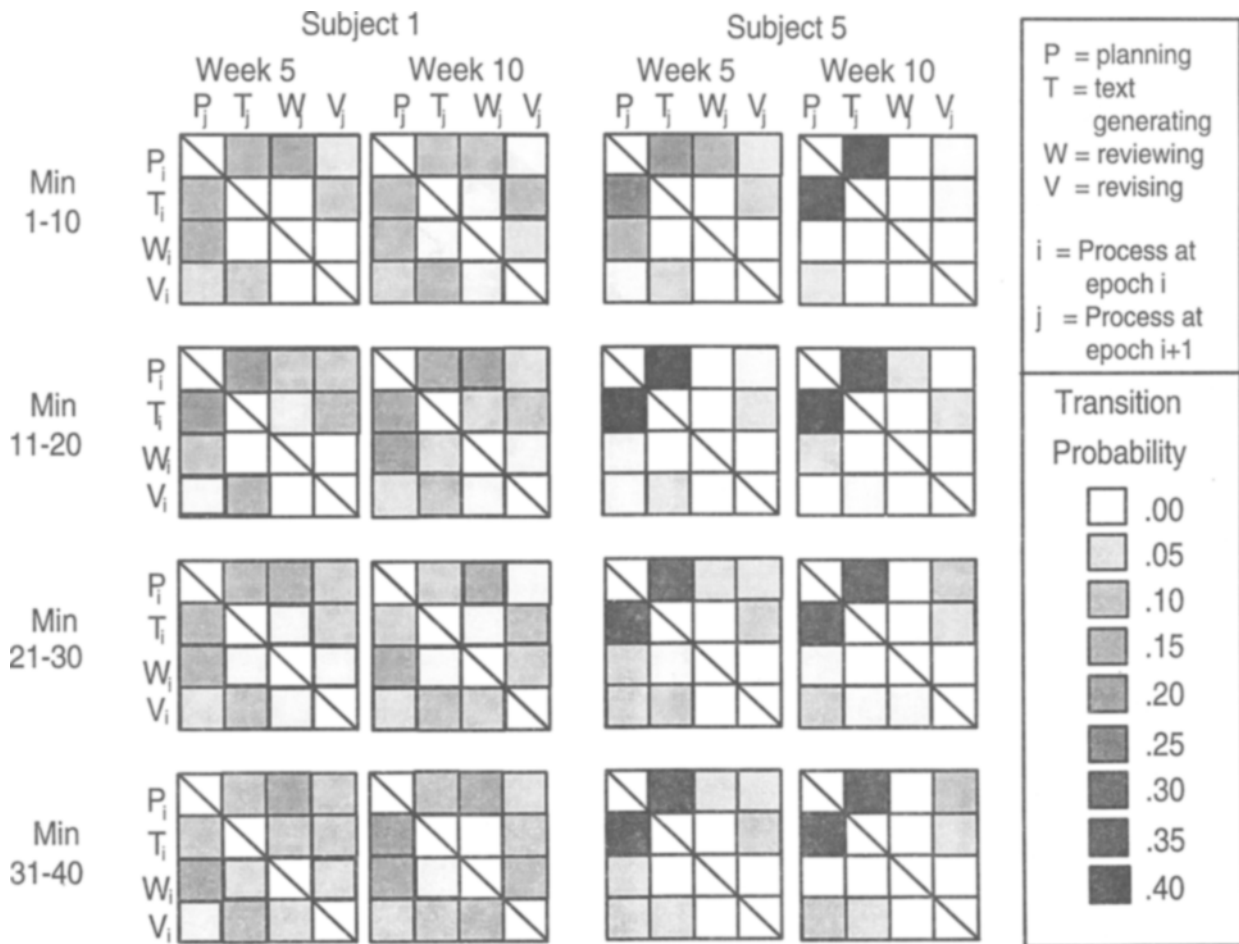
Figure 2. Transitional-probability matrices for Subjects 1 and 5 on Weeks 5 and 10 during four 10-min time blocks. Darker colors represent greater probabilities. The rows of each matrix represent the subprocess that occurred during Epoch i. The columns represent the subprocess that occurred during Epoch i+1. The main diagonal is undefined.

ities within each matrix is near 1.0, but is almost never unity because shifts from one of the four processes shown to "other" and from "other" to one of the four primary processes seldom sum to zero. As a result, the greatest probability that can be assigned to each off-diagonal cell is .083; the actual values vary across matrices and depend, as noted, upon the shifts to and from the "other" category. We computed chi-squares for each of the 236 complete matrices available for analysis; the remaining 4 matrices were unuseable owing to mechanical problems.[2] The analyses indicated that 98% of the matrices exhibited nonchance patterns ($p < .001$). Six of the writers' matrices reflected greater-than-chance shifts ($p < .001$) in processing 100% of the time. The other four writers' patterns were reliably different ($p < .001$) from chance 95% of the time or more.

To quantify the consistency that appeared in the transitional-probability matrices between and within writing sessions for individual subjects, we calculated correlation coefficients in two different ways. First, to measure the consistency across weeks, we treated the 12 defined elements in each transitional-probability matrix as a vec-

tor, and averaged these vectors across the four time blocks within each writing session. The average week-to-week correlations obtained in this way ranged from .56 to .85, and all were significantly greater than zero [i.e., $t(46) = 4.58$, $p < .001$]. The overall correlation across all subjects was .76. The correlation for Subject 1, whose data are shown on the left of Figure 2, was .59. For Subject 5, it was .94.

Within-session correlations were computed by averaging all of the vectors representing the matrices for Weeks 5–10 separately for each time block. These correlations ranged in value between .24 and .93, and averaged .74. The smallest r value was marginally significant ($z = 1.85$, $N = 60$, $p < .06$; the balance was reliability greater than zero). For Subject 1, the intrasession $r = 0.76$; for Subject 5, $r = 0.94$.

Quality profiles. The rated quality of our writers' documents was moderately high. The mean quality score (74% of the maximum possible points) was significantly greater than that produced by less proficient college students (57%) in one of our earlier studies (Ransdell & Levy, 1994), but less than that for professionally written passages excerpted

from *The New Yorker* magazine [82%; $F(2,114) = 8.26$, $MS_e = 31.5, p < .001$].

Quality scores in our sample ranged from 31%–98%. We tested for quality differences by comparing documents that fell in the highest and lowest quartiles of rated quality. The high-quality documents received an average score of 86% and the low-quality documents earned an average of 53% [$t(38) = 12.35, p < .001$]. Compared with low-quality documents, those rated high in quality contained more revising and reviewing epochs [40.5 vs. 28; $F(9,225) = 3.43, MS_e = 63.6, p < .001$], but about the same number of planning (46.2 vs. 45.5) and text-generating epochs (36 vs. 39.7). The essays rated highest in quality were produced by people who spent about 40% more time revising and reviewing than did those who produced documents scored as the low-quality essays [$F(1,29) = 5.54, MS_e = 818.5, p < .03$]. About 9.1% of each session was devoted to revision in the high-quality documents, compared with only 6.5% in the low-quality documents. These time differences were the result of more epochs being devoted to revising and reviewing rather than to longer epoch durations.

The total number of words created per minute (including those erased during editing and that did not appear in the final document) was significantly greater in high-quality documents (32.3 wpm) than it was in low-quality documents (24.7 wpm) [$t(35) = 2.24, p < .03$]. IRTs were greater during text generation for high-quality than for low-quality documents [177 msec vs. 52 msec; $t(37) = 2.30, p < .03$]. No other comparisons, including length of the final document, total words typed, percentage of words changed, or IRTs during planning, revision, or reviewing, were reliable.

Revision type. With practice generating protocols, writers made significantly more revisions, both meaningful and superficial. They made about 20 meaningful revisions per document starting at Week 4 and increased consistently to 33 revisions by Week 10 [$F(7,28) = 4.74, MS_e = 675, p < .001$]. Superficial revisions increased even more rapidly from about 20 at Week 4 to 67 at Week 10 [$F(7,28) = 3.04, MS_e = 1949, p < .01$]. These results are consistent with Ransdell and Levy (1994).

EXPERIMENT 2

The purpose of Experiment 2 is to compare writers' metacognitions about their writing with the actual time and effort they allocate to various writing subprocesses: Do writers allocate time and expend effort on the subprocesses of writing the way they think they do? We are particularly interested in whether writers changed their allocation strategies over time after they told us about how they think they write, and in the effects that generating verbal protocols might have on their metacognitions about their allocation of time and effort.

In a three-pronged approach to understanding writing (Trabasso & Suh, 1993), we seek converging evidence about the time and effort writers allocate to writing subprocesses. Metacognitions of the writing process would form part of the approach that includes formal models of writing such as the one proposed by Hayes and Flower (1980). Meta-

cognitions are not protocols, because they include writers' evaluations of how they believe they write. Verbal protocols expressly do not include evaluative comments (Ericsson & Simon, 1993). We compare three sources of data: writers' metacognitions about the time and effort they allocate to their writing processes, their verbal protocols, and their actual writing behavior over time.

Few studies have compared behavior and self-reports (Critchfield & Perone, 1990). In the domain of writing, we need to learn how behavior and self-report measures of the temporal qualities of composing correspond.

Method

Subjects. In a general psychology class at the University of Florida, 484 students completed a writing survey during their first week of the school term. Data from only 367 of these students were sufficiently complete to be useable. From this group, 40 volunteers were randomly selected to participate in an ostensibly unrelated writing study. A third group of subjects surveyed consisted of the 10 writers who participated in Experiment 1.

Materials. Two versions of a writing survey were used. A paper version inquired about subjects' writing skills and habits, such as how rapidly and accurately they could type, their word-processing experience, how often they wrote personal and business letters, the last time they wrote an "important" document (e.g., a term paper or college-admission essay), how their writing compared with their peers'. The critical elements of the instrument asked subjects to estimate the percentage of time that they spent planning and organizing before writing even began, writing, rereading what they wrote, pausing to think about what to write next, correcting spelling and grammatical errors, and adding ideas to and deleting ideas from what they wrote when they composed an "important document" using pen and paper, a typewriter, or a word processor.[3] (The few who could not type or had no word-processing experience were asked to imagine how they would divide their time.) Then they were asked to use a seven-point scale to indicate how effortful each of the writing subtasks was when the same type of document was composed using each of the three tools.

A software version of the questionnaire, developed for use in another project, was administered to the writers. This version was a superset of the paper instrument that probed more deeply into the subjects' writing experiences, beliefs, and impressions.

Procedure. During the first week of the school term, the surveys were administered during a class meeting. Those who indicated that they would be willing to participate in a two-part writing research project were later randomly selected until appointments were scheduled for 40 subjects. Their appointments for completing the questionnaires were for the fourth and sixteenth weeks of the semester, to coincide with the times during which the 10 writers were also surveyed. None of the subjects tested twice was given any reason to suspect that their last session in the study would involve a repetition of the writing survey. The subjects' final session was introduced as an opportunity to determine whether their writing patterns had changed during the intervening 12 weeks. These subjects did no writing in our laboratory. Their participation enabled us to establish a control baseline against which to compare the metacognitions of the writers who participated in Experiment 1.

Results and Discussion

The subjects' estimates of the percentage of time devoted to the various writing subtasks were added to yield estimates for planning, text generating, and revising/reviewing. The parallel survey items that involved estimates of mental effort were combined by averaging the ratings given to the appropriate items.

On the survey questions that dealt with estimated time and mental effort, for each subprocess there were no reliable differences between the responses given by the large class, the subset who reported for the repeated sessions, and the writers. At the outset, then, neither our selected writers nor our control subjects were atypical in their metacognitions about how they allocated their time and effort, compared with a heterogeneous group of undergraduates.

Figure 3 displays the estimated time on each writing subprocess for the writers and the subset from the large class and compares these with the average percentage of time that the writers devoted to the subprocesses during Weeks 4–10. Because the actual revising and reviewing times were so small (and generally not different from one another), in the subsequent analyses, revising and reviewing times were combined. Both groups of subjects initially believed that they devoted almost equal amounts of time to planning, text generating, and revising/reviewing; all between-group comparisons were nonsignificant. Their beliefs had not changed 12 weeks later.

Initially, writers significantly underestimated by about 58% the amount of time they spent on planning [$t(9) = 3.73$, $p < .001$], but their final estimates were in accord with the time that they devoted during Week 10 [$t(9) = 1.64$, n.s.]. The opposite pattern of results emerged for text generation: The difference between the pretest estimates and the actual time on Week 5 was not reliable, but their 72% posttest underestimation of their actual time on Week 10 was highly significant [$t(9) = 3.58$, $p < .005$].

The most dramatic finding, however, was the enormous discrepancy between estimated and actual time spent revising and reviewing. The writers initially estimated that they spent 33% of their time in these activities, but they actually spent about 5% on Week 5 [$t(9) = 13.78$, $p < .001$]. By the time of their posttest, the writers still believed that they spent about one-third of their time on these subprocesses, yet their actual time was only about 10% [$t(9) = 11.84$, $p < .001$]. Ten weeks of self-monitoring, a by-product of the talk-aloud requirement in Experiment 1, did very little to bring the writers' beliefs about their revising and reviewing times closer to reality or, for that matter, their behavior closer to their metacognitions.

Because the data using mental-effort estimates and the IRTs collected during Experiment 1 were measured on different scales, direct comparisons were achieved by dividing the range of IRTs by seven and assigning each IRT a value between one and seven. Figure 4 reveals a distinctly different pattern from that shown in Figure 3. Writers overestimated by 35%–42% their planning effort (IRT scores) both at the beginning of the experiment [$t(9) = 5.23$, $p < .001$] and at the end [$t(9) = 3.78$, $p < .001$]. Similar overestimations also occurred for text generation: the corresponding $t(9)$ values were 3.78 and 3.41, respectively (both ps $< .01$). In contrast, the writers' estimated effort during revising and reviewing closely resembled the adjusted IRT scores; all comparisons were nonsignificant.

An ANOVA indicated that the writers' estimated effort differed reliably among writing processes [$F(2,18) = 11.79$,
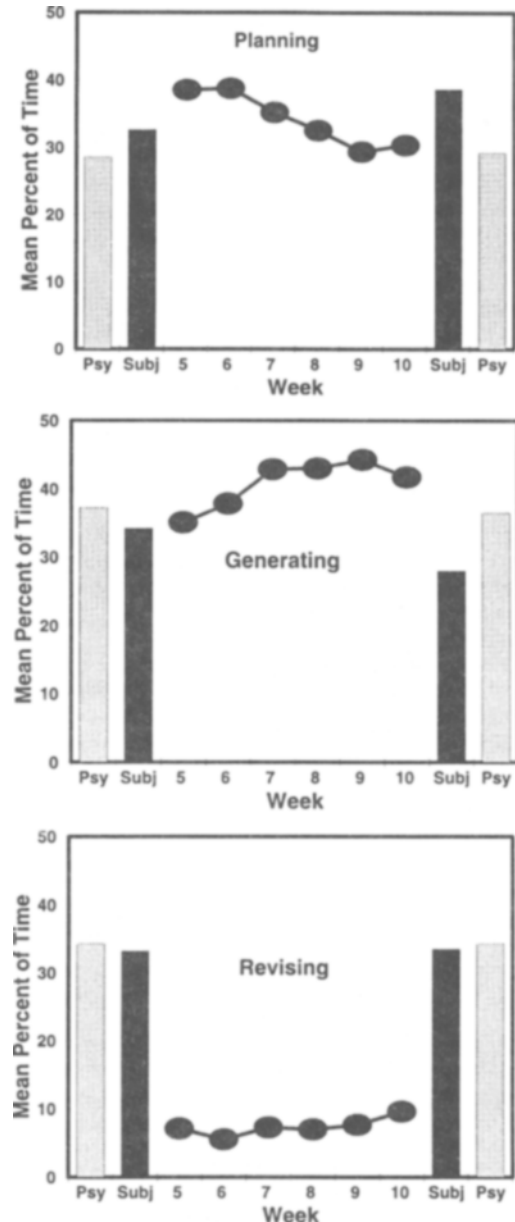

Figure 3. Mean percentage of time estimated during the pretests (left bars) and posttests (right bars) by the 10 writers (Subj) and the 40 control students from the general psychology class (Psy). The curves represent the actual percentage of time devoted to the writing processes across the 6 weeks during which the writers were required to generate verbal protocols.

$MS_e = 0.47$, $p < .001$]. Where writers indicated that they devoted greater-than-average effort to planning and text generation (about five on a seven-point scale), they indicated that their effort during revising and reviewing was about average (3.8). The adjusted IRTs, however, did not differ reliably across writing subprocesses. These estimates of their actual effort hovered around the midpoint of the scale.
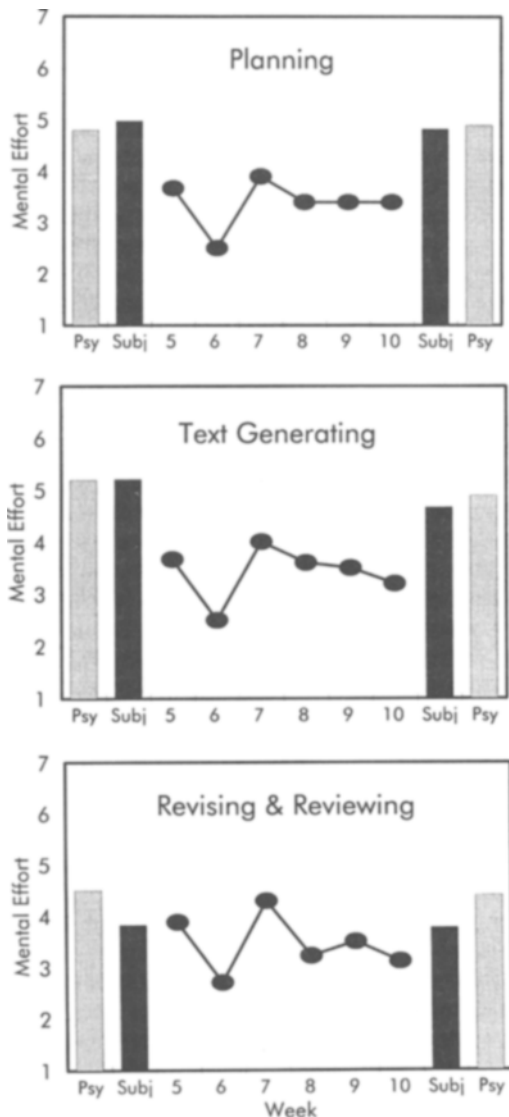
Figure 4. Mean estimated effort during the pretests (left bars) and posttests (right bars) by the 10 writers (Subj) and the 40 control students from the general psychology class (Psy). The curves represent the adjusted mental effort calculated from the IRTs (see text) for the writing processes across the 6 weeks during which the writers were required to generate verbal protocols.

GENERAL DISCUSSION

Our data suggest that studies of writing across time, while labor intensive, enable us to capture characteristics of composition that cannot be seen when subjects write only in a single session. The different allocation patterns for time and effort across weeks and the cumulative explanatory power of previous time expenditures on current efforts converge to suggest that writers can be distinguished by their processing profiles. These profiles, represented as transition-probability matrices, indicate a remarkable consistency of a writer's compositional strategies both within a specific writing session and across widely separate writ-

ing sessions. These profiles constitute a quantitative and graphical representation of a writer's composing *style*.

The time on task and effort data from Experiment 1 indicate that writers are not adversely affected by generating verbal protocols or by responding to a tone while writing. While these tasks make possible fine-grained analyses of writing subprocesses, they do not cause writers to significantly alter their time and effort strategies. According to Ericsson and Simon (1993), writing should not be affected by concurrent verbal protocols since they primarily involve the coding and decoding of verbal information. Previous research has found that the rate at which writers generate words is slowed while generating protocols (Ransdell, 1990). However, the present study indicates that this slowing of writing rate is at best transitory; writers are soon able to write efficiently even while generating protocols at the same time. Our data also clearly establish the independence of time and cognitive effort (indexed by IRTs) allocated to a writing subprocess.

In Experiment 2, writers were more accurate in predicting the amount of time and effort they allocate to planning and text-generation processes than they were in predicting that allocated to revising and reviewing. The subjects vastly overestimated time spent revising and reviewing even with practice generating verbal protocols. Part of the discrepancy between the self-reports in Experiment 2 and the data in Experiment 1 are reflected in the differences between Kellogg's (1987) directed retrospective time-allocation profiles and our own. For instance, in Kellogg's paradigm, subjects report higher levels of reviewing than we observed in our writing data, even when, as Kellogg does, we considered reviewing and revising as one process. Directed retrospection must partially reflect subjects' own "theories" because our subjects' metacognitions more closely match the writers' retrospections in Kellogg's studies than they do their own writing behavior. The fact that verbal protocols lack the characteristics of social communication, such as explanation, justification, and rationalization, sets them apart from self-report data (Ericsson & Simon, 1980, 1993). Self-reports are evaluative and often nonveridical. In contrast, verbal reports are nonevaluative and can predict patterns of actual behavior. Directed retrospection may fall between self-reports and verbal-protocol data in terms of matching actual writing behaviors.

While our writers were carefully selected, and generally did good work, we do not believe that these results are specific to relatively sophisticated writers. While some of our subjects were able to craft documents that surpassed the rated quality of published essays by professional writers, it is also true that they sometimes produced unquestionably inferior work. Because there were nowhere any differences between the initial metacognitions of our writers and those of the control subjects and the large heterogeneous group of subjects surveyed, our writers' understanding (and misunderstanding) of what they do when they compose may be representative of that of a diverse college population.

In the present experiments, writers were free to select the genre to use for any specific topic. A given writer some-

times wrote an essay, another time a letter or a dialogue. If it is plausible to hypothesize that these writers chose a specific genre for reasons that included their familiarity with its use, we should anticipate that their data will compare favorably with other situations in which the writer has experience with the genre called for by a situation—news styles for journalists, persuasive styles for lobbyists, highly formalized expository styles for scientific writers. It is an open empirical question as to whether they also apply to situations in which the writer is inexperienced with the genre demanded (e.g., in some composition classes) or has only a meager knowledge of the topic assigned.

There are many reports of the efficacy of planning in the preparation of high-quality documents. The writers in our study exhibited planning times approximating values reported by recent research in other laboratories (e.g., Kellogg, 1987). But planning time, as a global number measured across the writing session or as a fraction of the time spent within a temporal window as brief as 5 min, provides us no information that is useful in differentiating high-quality documents from low-quality ones. What did permit us to make such differentiations was the temporal distribution of time spent revising. Revising occurs throughout each writing session, albeit as a relatively small percentage of all writing activities. It is the revising that occurs late in a writing session, after a substantial amount of prose is developed, that contributes most to the differences among documents varying widely in quality.

The differences between people who produce exemplary writing and those whose writing is mediocre often can be attributed to differences in knowledge of the subject matter (Kellogg, 1987), in the genre (Ransdell, 1995; Ransdell & Levy, 1994), in the audience (Ransdell & Levy, 1994), or in mastery of writing conventions. These are macrolevel considerations that were largely held constant in this research. Our fine-grained examination of the interplay among writing subprocesses, of the extent to which the recursiveness that Emig (1971) first noted occurs, and of the robust and idiosyncratic writing signatures apparent in the writers' transitional probabilities may help us to understand better what contributes to differences in writing quality.

Some of our best and worst writing samples were composed by the same individuals. This situation provided a rare opportunity for inquiry because the variance attributable to individual differences was effectively reduced to zero. While the number of documents involved was only about 15% of the cases, the fact that such dramatic shifts in quality can occur within an individual must add an entirely new dimension to our conception of the writing process. A much larger sample of such cases is needed before it is possible to determine with any precision whether writers' transitional-probability signatures, distribution of time, or distribution of effort are the best predictors of the quality of their documents.

New models of writing will need to include production profiles of time and effort allocation that can predict quality writing. The present data suggest that those documents judged highest and lowest in quality can be differentiated by the number of times a writer initiates a revising epoch, re-

gardless of its duration. Quality may also be explained by the propensity of writers to oscillate rapidly between certain writing processes when they produce low-quality documents and a tendency to distribute their microwriting epochs (each only a few seconds in duration) more widely among writing processes when creating high-quality documents.

A test of Bereiter and Scardamalia's (1987) knowledge-transforming model of writing is beyond the scope of the present initiative, but evidence consistent with this model comes from the positive relationship between IRT and the quality of a document. Increased cognitive effort was particularly apparent during revising, indicating that the writers were not simply telling what they know, but were transforming their knowledge as well. It is possible, of course, that writers transform knowledge without an increase in cognitive effort, but the model suggests that such a relationship would be most likely under knowledge-transforming conditions.

A Time-and-Effort-Based Analysis of Writing

The time-and-effort-based analysis introduced here incorporates written, verbal, and resource-allocation protocols to account for writing behavior within and across writing sessions. This analysis adapts the three-pronged approach previously used by Trabasso and Suh (1993) in understanding text comprehension. Like the Hayes and Flower (1980) model of writing, it is a cognitive-process explanation, which focuses on a comparable set of writing subprocesses: planning, generating text, reviewing, and revising. What is new is that it incorporates time and effort components not apparent in research on verbal or written protocols alone. The analysis yields the following four empirically derived principles highlighted below: predictable patterns of process use occur within, and across, writing sessions, and positive relationships exist between text quality and effort and between text quality and time devoted to revision late in the writing episode.

Predictable Process Patterns Occur Within Writing Sessions

Time allocation is illustrated by SPLOMs (see Figure 1). These SPLOMs indicate, among other things, that writers trade off planning time for time generating text about two-thirds of the way into a given writing session. At this point, writers increase time generating text until about 85% of the writing session is complete. From there, text-generation time drops, possibly to allow time for increased revision. Revising accounts for only 2%–3% of writers' time until about halfway through a session, when it rises steadily to more than 10% by the end. Reviewing is similar in overall allocation, but begins to rise earlier, within the first quarter of each session.

Predictable Process Patterns Occur Across Writing Sessions

The analysis yields a pattern of time allocation to subprocesses that is stable across many sessions of writing. The pattern varies from one writer to the next in much the same way that cursive signatures vary. Like signatures, the writ-

ers' pattern of transition probabilities are remarkably stable through time. Future research should include a wider range of writing contexts to test the uniformity of this component. One strength of this analysis is that it incorporates the range in quality of writing that can occur within individuals, not simply the range that occurs between good and poor writers.

Effort Predicts Writing Quality

We found a positive relationship between IRT and the quality of a document that is consistent with Bereiter and Scardamalia's knowledge-transforming approach to writing and with Britton and Tessor's (1982) finding of greater cognitive effort among expert writers. Greater effort in higher quality essays does not support Kellogg's (1987) hypothesis, but in accord with that hypothesis, it is revising time that best predicts quality. Increased cognitive effort was particularly apparent during revising, indicating that the writers were not simply telling what they know, but were transforming their knowledge as well. It is of course possible that writers transform knowledge without an increase in cognitive effort, but our analyses suggest that such a relationship would be most likely under knowledge-transforming conditions.

Time Devoted to Revision Late in Writing Session Predicts Quality

The analysis relates these production profiles to a product-based marker of writing success—holistic quality. For example, it predicts that more time is allocated to revising in documents of relative high quality, especially within the last portion of a writing period. Self-report data indicate that writers believe they allocate upward of 30% of time to revision. Verbal protocols and actual writing behavior, in contrast, show that revision occurs at relatively low rates throughout (never more than 10% of total time). Such data are consistent with the observations of composition instructors that unskilled writers maintain that they devote substantial time to revising. Unskilled writers in fact characteristically treat revision simply as a matter of correcting superficial errors and spend very little time on this process. Our analysis reveals that revision takes less total time than other subprocesses, but that it disproportionately determines writing success. We also offer additional detail about the likelihood of writers changing from one process to another that might improve instructional techniques in composition. Our evidence suggests that higher quality documents differ not only in terms of allocation, but in terms of the number and distribution of subprocess epochs contained within writing processes.

One direct test of the utility of our approach will involve using time-and-effort-based process profiles from high-quality documents in long-term training sessions of people with varying experience and success in writing to determine whether their writing behavior can be trained to approximate such profiles. A stronger test would require that such modified profiles be associated with higher quality documents. The ultimate success of a time-and-

effort-based analysis of writing will be its ability to stimulate patterns of successful writing in a wide range of tasks.

Finally, it is clear from Experiment 2 that despite their misperceptions of the amount of time they devote to the various writing subprocesses, our subjects' understanding of how difficult writing is for them is quite different from their self-reports. Planning and text generating *seem* to require substantial effort (and more effort than reviewing and revising do), and this is generally interpreted by subjects as indicating difficulty. The actual estimates of effort or difficulty, as reflected by the adjusted IRTs, are consistent across both time and writing subprocess, but are generally less than the amount estimated. Only in the case of revising and reviewing, where the effort scores fall near the middle of the scale, are the subjects' metacognitions in accord with their behavior. Given that revising and reviewing occurred only about 10% of the time, we believe that greater weight should be given to the substantial overestimation of the effort required for planning and text generating. If that is done, the answer to the question about whether writing is really as difficult as it seems is: Not really.

REFERENCES

BEREITER, C., BURTIS, P. J., & SCARDAMALIA, M. (1988). Cognitive operations in constructing main points in written composition. *Journal of Memory & Language*, **27**, 261-278.

BEREITER, C., & SCARDAMALIA, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.

BRITTON, B. K., & TESSOR, A. (1982). Effects of prior knowledge on use of cognitive capacity on three complex cognitive tasks. *Journal of Verbal Learning & Verbal Behavior*, **5**, 421-436.

CRITCHFIELD, T. S., & PERONE, M. (1990). Verbal self-reports as a function of speed, accuracy, and reinforcement of the reported performance. *Psychological Record*, **40**, 541-554.

EMIG, J. (1971). *The composing process of twelfth graders*. Urbana, IL: National Council of Teachers of English.

ERICSSON, K. A., & SIMON, H. A. (1980). Verbal reports as data. *Psychological Review*, **87**, 215-251.

ERICSSON, K. A., & SIMON, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

ERICSSON, K. A., & SIMON, H. A. (1993). *Protocol analysis: Verbal reports as data* (rev. ed.). Cambridge, MA: MIT Press.

FODOR, J. D. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

GOULD, J. D. (1980). Experiments on composing letters: Some facts, some myths, and some observations. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 91-127). Hillsdale, NJ: Erlbaum.

HAYES, J. R., & FLOWER, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale, NJ: Erlbaum.

HAYES, J. R., & FLOWER, L. S. (1983). Uncovering cognitive processes in writing: An introduction to protocol analysis. In P. Mosenthal, L. Tamor, & S. A. Walmsley (Eds.), *Research on writing: Principles and methods*. New York: Longman.

HENDERSON, R. W. (1989). *Eventlog*. Iowa City: Conduit, University of Iowa.

KELLOGG, R. T. (1987). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory & Cognition*, **15**, 256-266.

KELLOGG, R. T. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 355-365.

KELLOGG, R. T., & MUELLER, S. (1993). Performance amplification

and process restructuring in computer-based writing. *Journal of Man-Machine Studies*, **39**, 33-49.

LEVY, C. M., FRYMAN, J. B., & RANSDELL, S. E. (1994). *Measuring response latency while running Windows applications programs*. Manuscript submitted for publication.

LEVY, C. M., & RANSDELL, S. [E.] (1994). Computer-aided protocol analysis of writing processes. *Behavior Research Methods, Instruments, & Computers*, **26**, 219-223.

LEVY, C. M., & RANSDELL, S. E. (in press). Writing signatures. In C. M. Levy & S. E. Ransdell (Eds.), *The science of writing*. Hillsdale, NJ: Erlbaum.

RANSDELL, S. E. (1990). Using a real-time replay of students' word processing to understand and promote better writing. *Behavior Research Methods, Instruments, & Computers*, **22**, 142-144.

RANSDELL, S. E. (1995). Generating thinking-aloud protocols: Impact on the narrative writing of college students. *American Journal of Psychology*, **108**, 89-98.

RANSDELL, S. E., & LEVY, C. M. (1994). Writing as process and product: The impact of tool, genre, audience knowledge and writer expertise. *Computers in Human Behavior*, **10**, 1-17.

RANSDELL, S. E., & LEVY, C. M. (in press). Working memory constraints on writing quality and fluency. In C. M. Levy & S. E. Ransdell (Eds.), *The science of writing*. Hillsdale, NJ: Erlbaum.

STRATMAN, J. F., & HAMP-LYONS, L. (1994). Reactivity in concurrent think-aloud protocols: Issues for research. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 89-112). Thousand Oaks, CA: Sage.

TRABASSO, T., & SUH, S. (1993). Understanding text: Achieving explanatory coherence through on-line inferences and mental operations in working memory. *Discourse Processes*, **16**, 3-34.

TYLER, S. W., HERTEL, P. T., McCALLUM, M. C., & ELLIS, H. C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning & Memory*, **5**, 607-617.

WILKINSON, L., HILL, M., WELNA, J. P., & BIRKENBEUEL, G. K. (1992). *Systat for Windows: Version 5* [Computer program]. Evanston, IL: Systat.

NOTES

1. Raters achieved high reliability by using the following procedures: (1) Initially, they read all of the sample quickly for scope. (2) They then read sets of 10 essays once through, evaluating each on a single quality dimension. They were instructed to determine first whether a document was high (5) or low (1) on the dimension. If they could not make this determination, they were to assign the document a rating of medium (3). Otherwise, they determined whether the document initially assigned a (5) justified that designation or whether a low-high (4) was warranted. Similarly, they determined whether those initially assigned a (1) warranted that value or a high-low (2). (3) This process was repeated for each of the 13 quality subscales. Raters met periodically to discuss the definitions of each subscale and to exchange examples outside the sample of essays that met each subscale criterion.

2. During Weeks 5–10, we calculated four transition-probability matrices for each subject during each writing session. Thus, 10 writers × 6 weeks × 4 time blocks = 240 matrices.

3. An "important document" was defined as one that had special significance to the writer (e.g., a letter to accompany a job application or a college-admission application). We assumed that targeting the editor of a national magazine as the audience for their documents and offering a $100 prize for the single best document would provide incentives for the subjects who composed in Experiment 1 to approach the task as one of writing an "important document."