ETS Research Report Series ISSN 2330-8516

Writing Processes in Short Written Responses to Questions Probing Prior Knowledge

Paul Deane, Tenaha O'Reilly, Szu-Fu Chao, and Kelsey Dreier

Educational Testing Service, Princeton, New Jersey

Author Draft

Abstract

The purpose of the report is to explore some of the mechanisms involved in the writing process. In particular, we examine students' process data (keystroke log analysis) to uncover how students approach a knowledge-telling task using 2 different task types. In the first task, students were asked to list as many words as possible related to a particular topic (word listing). In a second task, students were asked to write to a specific prompt that was designed to elicit their background knowledge of a topic using connected text (knowledge elicitation). Using a matrix incomplete block design, 1,592 high school students completed the 2 writing tasks in addition to a multiple-choice test of their background knowledge in 2 of 5 possible topics in the domain of U.S. history. An array of process data including students' typing and associated timing features was used to predict the writing scores on the 2 different types of tasks. The analyses revealed several distinct patterns that were associated with processing at the task knowledge productivity level, the editing effort level, and the keyboarding effort level. The robustness of the features was reflected in a set of hierarchal regressions that demonstrated that the process features were predictive of the writing score even when students' knowledge scores on the associated multiple-choice test were considered. In sum, the results indicate that process data in the form of log file analysis are useful for both understanding the writing process and exploring potential differences between students with high and low knowledge.

Keywords background knowledge, process data, keystroke analysis

Corresponding author: P. Deane, E-mail: pdeane@ets.org

Theories of writing typically postulate that increases in writing skills come as individuals increase the efficiency of specific processes and gain greater sophistication and control over the way those processes are coordinated. For instance, Hayes (2012) suggested four core writing processes: the *proposer* (which controls idea generation), the *translator* (which controls the conversion of ideas into specific language), the *transcriber* (which controls the process of converting language into actual text, through motor processes such as handwriting or keyboarding), and the *evaluator* (which monitors the other processes, evaluates progress, and may interrupt one process and give control to another). This basic model, shown in Figure 1, is in many ways skeletal, because it strips out the specific factors that differentiate stronger and weaker writers.

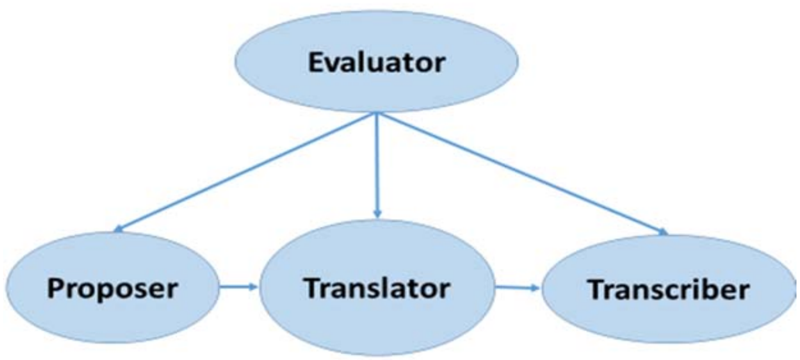

Figure 1 Component writing processes according to Hayes (2012).

The literature on how writing skill develops has emphasized the idea that core writing processes compete for working memory, and as a result, when one of the processes is inefficient and effortful (e.g., if a writer has difficulty with keyboarding or handwriting or has limited language skills), that reduces the capacity for other, critical writing tasks, such as idea generation, including advance planning (Kellogg, 2001; McCutchen, 1996; Olive, Kellogg, & Piolat, 2008). Developmentally, increases in writing skills are associated with transcription and language skills becoming increasingly automatized, combined with increased executive control that allows writers to apply appropriate writing strategies, depending on the writing task (Berninger, Winn, MacArthur, Graham, & Fitzgerald, 2006; Graham & Harris, 2000). From fairly early in the literature, this transition has been characterized as the shift from aknowledgetelling process characteristic of novice writers to a knowledge-transforming process characteristic of expert writers (Bereiter & Scardamalia, 1987). In a knowledge-telling approach, the writer retrieves information from memory and puts that information into written words, with minimal executive control or evaluation. In what Bereiter and Scardamalia (1987) characterized as a knowledge-transforming approach, more sophisticated strategies are employed that result in a more recursive process, in which the evaluator may reject or revise ideas that have already been expressed or the way those ideas have been phrased, resulting in a cyclical process in which drafting, revision, and editing processes happen not as a purely sequential series of steps but interleaved in time.

However, there are contexts in which it is clearly appropriate for writers to follow a knowledge-telling strategy. For example, if asked a factual question to which one knows the answer, the most appropriate strategy is to search for that information in memory, retrieve it, and express it in words, without significant effort devoted to revising or editing the resulting text. So-called knowledge-transforming strategies require significant effort, and the application of sophisticated evaluation and revision strategies is not always necessary and may even be counterproductive in contexts where speed and efficiency of processing are at a premium. Part of writing skills is the ability to choose a strategy appropriate to the task at hand (Breetvelt, Van den Bergh, & Rijlaarsdam, 1994; Snow, Allen, Jacovina, Perret, & McNamara, 2015), which in some contexts may be a simple knowledge-telling strategy in which revision and editing play a minimal role.

Keystroke Log Analysis

Keystroke log analysis provides a method to observe temporal patterns of text production in real time, which support inferences about the kinds of writing processes and strategies writers use to accomplish a task (Kaufer, Hayes, & Flower, 1986; Miller, 2000; Van Waes, Leijten, Wengelin, & Lindgren, 2011; Wengelin, 2006). Developmental increases in the automaticity of text production can be measured using this technique (Alves, Branco, Castro, & Olive, 2012; Alves & Limpo, 2015).

The literature has suggested significant associations between writing processes and writing quality (Connelly, Dockrell, Walter, & Critten, 2012; Kaufer et al., 1986). For stronger writers, text tends to be produced efficiently in longer bursts; pauses are more likely to happen at natural loci for planning, such as at clause and sentence boundaries; and more editing and revision behavior can be observed. Among weaker writers, text tends to be produced less efficiently and pauses appear in locations that suggest difficulties in typing, spelling, word finding, and other transcription processes. We have obtained similar results in analyses of keystroke patterns among middle school students completing various essay tasks (Deane, 2014; Deane & Zhang, 2015; Zhang & Deane, 2015).

The studies we cite share one striking feature: Nearly all of the features that correlate with writing quality can be interpreted in terms of a simple, knowledge-telling approach to writing in which students generate ideas and then write them down in a simple, sequential fashion without recourse to complex revision or editing strategies. Deane (2014) and Zhang and Deane (2015) observed that the strongest associations with essay scores are factors that appear to measure general fluency and sentence-level planning and deliberation. Deane and Zhang (2015) observed that the best predictors of writing quality are total time on task, length of bursts of text production, duration of pauses within and between words, and (with relatively low weight) the duration of simple editing events, such as cutting/pasting and backspacing. Except for the editing events, these are features that measure the fluency with which writers can generate ideas, translate them into words, and output them using a keyboard. This result can be explained easily if we assume that the writers in these studies primarily applied a knowledge-telling strategy to the essay-writing task. On this interpretation, the writers in these studies typically produced their essays sentence by sentence and paragraph by paragraph, more or less sequentially with relatively little editing, and they differed primarily in how easily they were able to generate the content. This interpretation is reasonable because the populations examined in the cited studies are mostly drawn from middle school and lower grade levels, where we can expect fluency in basic writing processes to be a significant differentiator between students with higher and lower performance.

However, this interpretation raises important questions about how features derived from a keystroke log should be interpreted and the extent to which feature properties and their interpretations may shift from one kind of writing task to the next. Writing tasks differ in their complexity, amount of time required, the evaluation standards to be applied, and in many other ways. All of these affect both the difficulty of the writing tasks and the kinds of writing strategies that writers will (or should) apply to these. However, little work has been done to examine the ways in which the dynamics of writing behavior shift from one kind of task to the next or to evaluate how the interpretation of keystroke log data may be affected by parameters on which writing tasks may differ.

The purpose of this study is to begin to lay the foundations for this kind of analysis by considering a special case: a very simple writing task in which a knowledge-telling strategy is not only effective but entirely construct appropriate, namely, short-answer questions designed to probe background knowledge about a specific topic. In this kind of question, answers are only evaluated with respect to the accuracy with which they reproduce the correct content, and there should be little occasion for writers to engage in revision or editing behaviors, except for local error correction. As a result, writing behavior on this kind of task is likely to represent an almost pure example of knowledge telling, in which the primary drivers of task difficulty are the ease with which the writer can retrieve relevant information and the fluency with which he or she can express it.

Research Questions

Our goal in this study was to examine whether, when students are given a pure knowledge-telling task, they display similar behavior to that we have observed in middle school student essays (which we have interpreted as representing a dominant knowledge-telling strategy). In particular, we would like to explore the following specific research questions (RQs):

RQ 1. Do students respond to short-answer background-knowledge questions primarily by appending information sequentially to their existing answers?

RQ 2. When students respond to short-answer background-knowledge questions, what factor structure is associated with keystroke features derived from student responses? How interpretable are the resulting factors?

RQ 3. Which of the factors identified to answer RQ 2 are predictive of high performance? How much unique variance do they account for?

RQ 4. Are the values of keystroke features sensitive to the differences between different kinds of short-answer questions? Can observed differences be accounted for in terms of construct differences between the item types?

We intended to provide a baseline set of results for studying knowledge telling as a strategy and to compare those results to those obtained in prior studies focused on student essay writing. It is of interest to clarify which features of student essay writing are paralleled in tasks for which a pure knowledge-telling strategy is appropriate by definition.

Method

Participants

Participants were 1,592 9th- to 12th-grade students from five U.S. schools. Schools received a payment prorated according to the number of participating students. This sample contained slightly more individuals self-identified as male than as female (803 vs. 707, with 82 not indicated); 36.6% self-identified as Latino, 27.7% as White, 4.6% as Asian or Asian American, 3.2% as Black or African American, 1.7% as American Indian or Alaskan Native, 1.4% as native Hawaiian or Pacific Islander, and 2.5 % as other, whereas 13.6% indicated a mixed background (more than one category selected), 6.6% indicated that they preferred not to answer, and 2.3% made no selection.

Materials

Each student was administered one of four forms, each focused on background knowledge in a specific history topic. Two topics, immigration (Form 1) and women's suffrage (Form 4), were based on texts that appeared in two GISA[1] forms (O'Reilly, Weeks, Sabatini, Halderman, & Steinberg, 2014). Two additional topics, the American Civil War (Form 2) and colonial America (Form 3), were chosen due to their consistent inclusion in high school curricula and textbooks. The topics additionally represented varying chronologies within the domain of history, ranging from 17th-century colonial America to women's suffrage in the 20th century. A final, fifth topic was developed, general U.S. history, containing broad questions throughout all of U.S. history, though it is not included in the current analysis.

Multiple-Choice Questions

Each form included a battery of multiple-choice items testing student knowledge of a specific history topic, designed to cover both basic and conceptual knowledge about the topic. These items were developed as follows. After topics were selected, questions were first pulled from the background section of existing GISA forms (O'Reilly et al., 2014) and from the 8th- and 12th-grade National Assessment of Education Progress 2014 U.S. history assessments (National Center for Education Statistics, 2014). Next, published state tests from various regions of the United States were used to find additional questions. Each form contained a small set of anchor items designed to link performance across forms (which are excluded from the present analysis).

In the present study, we used total raw score on the multiple-choice questions as an additional measure of knowledge about the specific topic tested in each form. There were a total of 50 multiple-choice questions on immigration (Form 1), 63 on the Civil War (Form 4), 59 on colonial America (Form 3), and 50 on women's suffrage (Form 2). Reliabilities for the multiple-choice question sets were between .68 and .82 (α = .71 on Form 1, .68 on Form 2, .82 on Form 3, and .72 on Form 4).

Constructed-Response Questions

Each of the four forms also included two types of constructed-response items. The first type, which we term *word listing* items, identified a topic and asked the students to list as many words as they could that were closely related to that topic. Students had 2 minutes to respond. The second type, which we term *knowledge elicitation* items, identified a topic and asked students to write what they knew about a specific aspect of that topic. Students had 5 minutes to respond. The specific topics tested on each form and the wording of each question are shown in Table 1. At the end of the test, after the constructed-response items and the multiple-choice items had been completed, students had the opportunity to revise their answers to the word listing item. The revision gave an opportunity to measure whether exposure to the multiple-choice questions primed students' awareness of and possibly increased their knowledge about the topic. While the revised word listing item is of secondary interest in this study, we include it in parts of the analysis as a cross-check on the results from the initial word listing item. That is, we expect the correlation between multiple-choice items and the word listing item to increase after revision.

Table 1 Forms and Questions Administered

Form	Word listing question	Knowledge elicitation question
1	Please list as many words as you can related to **United States Immigration** in the box below.	Tell us what you know or have learned about **diversity in America**. Please explain what it is and how it is related to the study of **19th century United States Immigration** more generally. Please use as many specific terms related to the topic and specific examples as you can.
2	Please list as many words as you can related to the **women's right to vote** in the box below.	Tell us what you know or have learned about the **right to vote**. Please explain what it is and how it is related to the study of the **Women's Rights Movement** more generally. Please use as many specific terms related to the topic and specific examples as you can.

3	Please list as many words as you can related to **Colonial America i**n the box below.	Tell us what you know or have learned about the **thirteen original American colonies**. Please explain what they are and how they are related to the study of **Colonial America** more generally. Please use as many specific terms related to the topic and specific examples as you can.
4	Please list as many words as you can related to the **American Civil War** in the box below.	Tell us what you know or have learned about the **Emancipation Proclamation**. Please explain what it is and how it is related to the study of the **American Civil War** more generally. Please use as many specific terms related to the topic and specific examples as you can.

Procedure

Each student was assigned to complete one of the four forms during a single class period in his or her social studies class. All forms were administered digitally, using an in-house test delivery platform that logged the students' writing processes and collected their final answers to each question. Parental permissions were obtained, responses checked to make certain that only responses with appropriate permissions were included in the analysis, and personally identifiable information was removed. A total of 402 students completed both constructed-response items in Form 1 (out of 423 total). A total of 399 students completed both constructed-response items in Form 2 (out of 399 total). A total of 389 students completed both constructed-response items in Form 3 (out of 399 total). A total of 401 students completed both constructed-response items in Form 4 (out of 402 total).

Scoring

The keywords present in student responses were analyzed using in-house topic analysis tools to identify groups of words that were relevant to the assigned topic. Individual keywords (where central to the assigned topic) and groups of keywords (where less central) were assigned weights. The least weight was assigned to words from word families that were present in the question prompt. More weight was assigned to words that were somewhat or very relevant to the assigned topic. The most weight was assigned to words that indicated detailed and specific knowledge about the assigned topic, which (in the case of knowledge elicitation items) specifically included information about the focus identified in the prompt. For example, students got minimal credit on the knowledge elicitation question on Form 2 for specifically mentioning women, the right to vote, or the Women's Rights Movement. They obtained some credit for mentioning specific relevant topics, including words related to voting, constitutional amendments, and protests, but got the most credit for mentioning such topics as feminism and specific terms and names, such as suffrage, the Seneca Falls convention, the 19th Amendment, Elizabeth Cady Stanton, or Susan B. Anthony. The weights assigned to specific keywords and keyword groups were adjusted until the sum of the weights produced a sorting that assigned top scores to obviously strong responses and only included weak responses (with no relevant keywords, except possibly prompt words) at the lowest score levels.

Planned Data Analysis

We conducted four major analyses, corresponding to our four research questions.

Distribution of Keystroke Events

To describe general patterns of student behavior, and therefore answer RQ 1, we extracted the following features.

- *Proportion of time spent appending to the end of the response (proportion time after last character)*. We hypothesized that students are applying a knowledge-telling strategy in which they recall information and then write down whatever they have recalled and therefore spend almost all their time appending new information at the ends of their current responses.
- *Distribution of keystroke events.* Keystroke events were classified into five event types: insert, backspace, delete/cut (delete one or more characters without backspacing), paste (insert three or more characters simultaneously), and replace (insert and delete characters simultaneously). We hypothesized that the vast majority of keystroke events would be insert events, supplemented with backspace events (corresponding to purely local edits). Delete/cut, paste, and replace correspond to revision activities that should be very infrequent in a task that requires a knowledge-telling strategy.

If students are writing in a knowledge-telling mode, in which they produce text in the order in which they recall information about the topic, we would expect the overwhelming majority of events in the keystroke log to occur at the end of the text and for most events in the log to consist of insertion events, with occasional backspacing to correct spelling errors and typos.

Features Used to Characterize Writing Patterns

To identify the factors we expected to find in student responses, and therefore answer RQ 2, we extracted several features from student responses for the purpose of entering them into a factor analysis. These features are a subset of those available from our larger keystroke analysis system and were selected to satisfy three criteria, except where noted.

1. Where possible, we selected features where there was a significant correlation greater than .2 between the feature value obtained for this feature in the word listing item and the corresponding feature value for the knowledge elicitation item in at least one of the four forms. Many of the features identified in this way were in the initial set of features described by Almond, Deane, Quinlan, and Wagner (2012).
2. The features could be defined meaningfully without missing data, even for behaviors like editing, in which the data were very sparse. We thus favored features for which we could provide meaningful default values for feature calculation.
3. If possible, the feature was related to one of the factors we identified in previous work (Deane, 2014; Deane & Zhang, 2015; Zhang & Deane, 2015), in particular, planning and deliberation, fluency (or keyboarding effort), and effort put into local editing or revision. However, given the first two constraints, we did not use feature sets identical to those reported in Deane (2014), Deane and Zhang (2015), or Zhang and Deane (2015).
4. Variables were transformed to approximate a normal distribution when possible. To this end, count features were transformed by taking the square root and duration and relative probability features by taking the log.

The following features were selected to provide evidence about the overall tempo of text production (corresponding, roughly, to the natural unit for planning a phrase or a clause). We defined *phrasal bursts* by dividing a writer's response into groups of words produced together, using long pauses (defined as any pause between words four times longer than that individual's median interkey pause) to identify the beginnings and ends of bursts. Given this definition, we then recorded the following information.

- *The maximum length of phrasal bursts (in words; maximum phrasal burst length).* Students who can produce longer sequences of words without stopping to plan presumably are able to access the knowledge they are expressing more readily and are better able to express and transcribe that knowledge efficiently.

- *Total time spent pausing at the start of a phrasal burst (total time at burst start).* The pause time before a phrasal burst is presumably being used to plan the text to be produced.
- *Total time spent pausing between phrasal bursts (total time between bursts).* If there are multiple long pauses before a burst, with minimal text inserted in between, these additional pauses probably also represent time the writer needed to plan the text to be produced.

The following features were selected to provide evidence about the tempo of text production at a somewhat finer grain size, using a slightly different definition of *burst*, which we will term *fast bursts*, in which any pause longer than two thirds of a second was treated as defining the end of a burst. We reasoned that this cut point (which is long enough to exclude most keystrokes produced while typing individual words) would provide a window into students' lower level keyboarding and language skills. If students can type a word or sequence of words without any pauses longer than two thirds of a second, most of the effort they are making is likely to focus on spelling and transcription processes. In particular, we recorded the following information.

- *Number of fast bursts*. The number of bursts produced can be viewed as a measure of overall fluency of idea generation. Students who produce more fast bursts are likely to produce more text, though there may also be a trade-off between burst length and the number of bursts.
- *Mean log length of fast bursts in characters (fast burst length).* The length of a fast burst in characters is likely to be related to the speed of keyboarding. It is the number of characters the writer can produce quickly without generating fresh content.

The following measures were also intended to measure fluency of text production by providing average measures of latency and speed.

- *Mean log duration of keystroke production events (base text production latency; excluding only text positions likely to involve word finding or editing, such as the first character in a word or the pause before a cut, paste, or jump event).* We hypothesized that the number of keystrokes per second that a writer can produce, including backspaces, but excluding such pauses, would primarily reflect keyboarding skill.

- *Mean log duration of a single keystroke within a word (duration of within-word pauses).* We hypothesized that the latency before characters are inserted within a word would provide a slightly different measure of fluency of text production, because in-word pauses are likely primarily to reflect keyboarding effort.
- *Mean log duration of a single keystroke between words (duration of between-word pauses).* We hypothesized that this feature would provide a measure of the extent to which pauses outside a word may be affected by other kinds of writing processes, such as lexical access and sentence-level planning.

The following measures were intended to capture pause behaviors likely to reflect effort in text production or editing effort. Most of these features were defined as some kind of log odds calculation, comparing more effortful behavior (such as the pause before a word, the pause before an edit, or the pause before a backspace action) to less effortful behavior (such as the pause before inserting another character inside a word). In particular, we recorded the following information.

- *The relative balance between active writing time and other activities (editing, planning; log ratio productive to nonproductive time).* We took the time spent on ordinary text production (inserting characters in a word, on whitespace or punctuation marks between words, but excluding word-initial pauses, backspacing, editing, or pauses between sentences) and compared it to the total time spent in all those other activities. We hypothesized that the log of this ratio would provide a measure of the relative effort the writer put into text production compared to all other writing processes.
- *The relative length of time spent inactive before beginning to write (relative start time).* We took the total time a writer spent before beginning to type and compared it to the total time a writer spent inserting characters within a word. We hypothesized that the log of this ratio would provide a measure of the relative effort the writer put into planning or generating ideas before text production began.
- *The relative length of time spent producing the first characters of words (relative word-initial time).* We took the total time a writer spent producing the first character of a word and compared it to the total time a writer spent inserting characters within a word. We hypothesized that the log of this ratio would provide a measure of the

relative effort the writer put into such processes as sentence-level planning, lexical access, or spelling.

- *The speed with which the first characters of words are individually produced (word start speed).* This is a simple rate: time spent producing first characters of words divided by the number of such characters produced.
- *The relative length of time spent backspacing over existing text (relative backspacing time).* We took the total time a writer spent backspacing and compared it to the total time a writer spent inserting characters within a word. We hypothesized that the log of this ratio would provide a measure of the relative effort the writer put into monitoring his or her writing for typographic errors.
- *The relative number of characters deleted versus inserted inside a word (log odds of deletion).* We took the number of characters inserted and compared it to the number of characters deleted inside a word. We hypothesized that the log of this ratio would provide a measure of the balance between insertion and deletion in the text.
- *The efficiency of text production (keystroke efficiency), as measured by the ratio between the final number of characters produced and the total number of characters inserted or deleted.* If a writer showed hesitancy or inefficiency in text production, the writer might produce and then delete large amounts of text that never made it into the final written product. We hypothesized that this proportion would provide an alternative measure of editing behaviors.

To make these calculations insensitive to missing data, we added a number equivalent to the smallest possible value (.01 seconds for durations and 1 second for counts) to each value entered in the comparison, before taking the logarithm.

Hierarchical Linear Regression Against Item Scores

To address RQ 3, we calculated correlations among item scores and estimated factor scores, conducted hierarchical multiple linear regressions using item scores as the independent variable, and estimated factor scores (plus multiple-choice total scores) as the dependent variables to determine how much unique variance in scores each variable accounted for.

Comparison of Means

Finally, to address RQ 4, and evaluate the effect that item type had on keystroke values, we conducted paired-sample *t*-tests to see whether the distribution of feature values differed in word listing and knowledge elicitation items for each feature type. The total time and number of word features were excluded, because there was little need to confirm that students would spend more time, and produce more words, in the knowledge elicitation task.

Data Preparation and Screening

All features were calculated twice, once for students' responses to the word listing task and once for their responses to the knowledge elicitation task, and treated as separate variables in the output. Because missing data primarily occurred where a particular type of keystroke event did not occur in an individual log (e.g., where respondents did not produce any backspacing sequences), these data were missing in the student's behavior. We therefore excluded missing values pairwise. The required minimum sample size for factor analysis was satisfied as described in Child (2006), with final samples sizes of 424 for Form 1, 399 for Form 2, 389 for Form 3, and 402 for Form 4, yielding between 18 and 19 cases per variable.

Results

Distribution of Keystroke Events

Students consistently spent the overwhelming majority of their time appending to the end of the responses. For all eight items, students spent between 97% and 98% of their time producing text sequentially as opposed to making major sentence or section-level revisions (which would be suggestive of an iterative and evaluative knowledge-transforming approach; see Table 2). More precisely, the vast majority of students' time was spent either inserting new text or making only minor edits at the point they are inserting new text, such as by backspacing over immediately adjacent text (see Table 3). Such behavior is consistent with a knowledge-telling approach.

Table 2 Proportion of Time Spent After Last Alphanumeric Character, by Item, in Percentages

Form	Mean	*SD*
Form 1		
Word listing	98	6
Knowledge elicitation	98	6
Form 2		

Word listing	98	8
Knowledge elicitation	98	6
Form 3		
Word listing	98	6
Knowledge elicitation	97	6
Form 4		
Word listing	98	7
Knowledge elicitation	97	8

Table 3 Distribution of Keystroke Events by Type, in Percentages

Form	Insert	Backspace	Delete/cut	Paste	Replace
Form 1					
Word listing	85.5	13.5	0.7	0.1	0.2
Knowledge elicitation	86.5	12.6	0.6	0.1	0.2
Form 2					
Word listing	87.7	11.4	0.4	0.1	0.3
Knowledge elicitation	87.7	11.4	0.5	0.1	0.3
Form 3					
Word listing	86.4	12.8	0.5	0.1	0.3
Knowledge elicitation	86.9	12.3	0.5	0.1	0.3
Form 4					
Word listing	85.6	13.4	0.7	0.1	0.1
Knowledge elicitation	87.6	11.6	0.4	0.1	0.3

Factor Analysis

Exploratory factor analysis was performed to identify latent factors that appeared to underlie writing behavior on the word listing and knowledge elicitation tasks. We calculated factor analyses separately by form and prompt. For each form, we thus obtained two sets of factors (one for the word listing item and one for the knowledge elicitation item).

Most of the features initially entered into the analysis met several of the standard criteria for factorability. In particular, these criteria and decision points included the following elements:

- In the four forms, nearly all variables had at least one correlation at or above .3. The only exceptions were total time at phrasal burst start on the knowledge elicitation items in Form 1, with the strongest correlation at .26, and the word listing items in Forms 2 and 4, with the strongest correlations at .26 and .29.

- The Kaiser–Meyer–Olkin sampling adequacy measures for the four forms were .77 for Form 1, word listing; .76 for Form 1, knowledge elicitation; .72 for Form 2, word listing; .73 for Form 2, knowledge elicitation; .79 for Form 3, word listing; .77 for Form 3, knowledge elicitation; .71 for Form 4, word listing; and .74 for Form 4, knowledge elicitation.
- Finally, all communalities were above .3, except for (a) two forms for total time between phrasal bursts, in which the communalities were .19 in Form 1, word listing, and .29 in Form 3, knowledge elicitation, and (b) five of the eight forms for total time at phrasal burst start, in which the communalities were .15 for Form 2, word listing; and Form 4, knowledge elicitation; .25 for Form 1, knowledge elicitation, and Form 3, word listing; and .26 for Form 4, word listing.

Overall, these statistics indicated that it was appropriate to conduct factor analysis with these features, although the two time features appeared to be somewhat weaker in their shared variance than desirable in many forms.

We specifically applied principal axis factoring to identify potential latent variables that accounted for common variance across features. Very similar results were observed on all items.

- On Form 1, word listing, the first factor accounted for 33.8% of the variance, the second factor for 14.9% of the variance, the third factor for 14.1% of the variance, and the fourth factor for 8.1% of the variance (with an eigenvalue just over 1).
- On Form 1, knowledge elicitation, the first factor accounted for 35.0% of the variance, the second factor for 17.2% of the variance, the third factor for 13.7% of the variance, and the fourth factor for 8.7% of the variance (with an eigenvalue just over 1).
- On Form 2, word listing, the first factor accounted for 31.4% of the variance, the second factor for 15.6% of the variance, the third factor for 13.0% of the variance, and the fourth factor for 10.4% of the variance (with an eigenvalue just over 1).
- On Form 2, knowledge elicitation, the first factor accounted for 30.8% of the variance, the second factor for 20.6% of the variance, the third factor for 12.0% of the

variance, and the fourth factor for 10.1% of the variance (with an eigenvalue just over 1).

- On Form 3, word listing, the first factor accounted for 35.1% of the variance, the second factor for 18.7% of the variance, the third factor for 9.1% of the variance, and the fourth factor for 8.8% of the variance (with an eigenvalue just over 1).
- On Form 3, knowledge elicitation, the first factor accounted for 37.4% of the variance, the second factor for 16.9% of the variance, the third factor for 9.4% of the variance, and the fourth factor for 894% of the variance (with an eigenvalue just over 1).
- On Form 4, word listing, the first factor accounted for 32.3% of the variance, the second factor for 17.0% of the variance, the third factor for 13.4% of the variance, and the fourth factor for 9.1% of the variance (with an eigenvalue just over 1).
- On Form 4, knowledge elicitation, the first factor accounted for 34.4% of the variance, the second factor for 17.9% of the variance, the third factor for 10.2% of the variance, the fourth factor for 9.3% of the variance, and the fifth factor for 6.8% of the variance (with an eigenvalue just over 1).

Four-factor solutions emerged with a cutoff eigenvalue of 1 for seven of the eight items. We therefore preferred a four-factor solution for all items, using Promax rotations of the factor loading matrix, because we expected that there would be correlated factors. These solutions accounted for 70.9% of the variance in Form 1, word listing; 74.5% of the variance in Form 1, knowledge elicitation; 70.5% of the variance in Form 2, word listing; 73.4% of the variance in Form 2, knowledge elicitation; 71.7% of the variance in Form 3, word listing; 72.6% of the variance in Form 3, knowledge elicitation; 71.8% of the variance in Form 4, word listing; and 71.8% of the variance in Form 4, knowledge elicitation. The final set of solutions, extracting four parallel factors for each form, are shown in Tables 4–11.

Table 4 Rotated Factor Loadings (Pattern Matrix) and Communalities for the Form 1 Word Listing Item, Based on a Factor Analysis With Promax Rotation for 15 Process Log Features

Process Log Feature	Production effort	Keyboarding effort	Editing effort	Overall deliberation time	Communalities
Relative word-initial time	1.00	−0.23		0.22	0.91
Relative productive time	−0.95				0.82
Relative start time	0.89				0.83
Word start speed	0.77				0.67
Number of fast bursts	−0.70			0.35	0.80
Duration of between-word pauses	0.55	0.34			0.54
Maximum phrasal burst length	−0.48	−0.26			0.35
Duration of within-word pauses		0.78			0.58
Fast burst length		−0.77			0.61
Base text production latency		0.54			0.31
Log odds of backspacing actions			0.97		0.85
Relative backspacing time			0.85		0.72
Keystroke efficiency	0.40		−0.47		0.54
Total time at phrasal burst start				0.67	0.40
Total time between phrasal bursts	0.25			0.33	0.19

Note. Factor loadings < .2 are suppressed.

Table 5 Rotated Factor Loadings (Pattern Matrix) and Communalities for the Form 1 Knowledge Elicitation Item, Based on a Factor Analysis With Promax Rotation for 15 Process Log Features

Process Log Feature	Production effort	Keyboarding effort	Editing effort	Overall deliberation time	Communalities
Relative word-initial time	0.79		0.22	0.24	0.79
Relative productive time	−0.91				0.74

Relative start time	0.91				0.83
Word start speed	0.48			0.40	0.66
Number of fast bursts	−0.96			0.26	0.83
Duration of between-word pauses	0.32	0.52		0.35	0.83
Maximum phrasal burst length	−0.64	−0.31			0.50
Duration of within-word pauses		0.93		−0.42	0.81
Fast burst length		−0.75			0.60
Base text production latency		0.62			0.47
Log odds of backspacing actions	0.27		0.85		0.69
Relative backspacing time			0.81		0.73
Keystroke efficiency	0.26		−0.58	0.39	0.68
Total time at phrasal burst start	−0.48			0.42	0.25
Total time between phrasal bursts				0.69	0.42

Note. Factor loadings < .2 are suppressed.

Table 6 Rotated Factor Loadings (Pattern Matrix) and Communalities for the Form 2 Word Listing Item, Based on a Factor Analysis With Promax Rotation for 15 Process Log Features

Process Log feature	Production effort	Keyboarding effort	Editing effort	Overall deliberation time	Communalities
Relative word-initial time	1.05			0.21	0.95
Relative productive time	−1.01	0.21		−0.24	0.91
Relative start time	0.80			−0.30	0.87
Word start speed	0.79				0.56
Duration of between-word pauses	0.51	0.39			0.46
Maximum phrasal burst length	−0.40	−0.21			0.33
Duration of within-word pauses		0.73			0.52
Fast burst length		−0.73			0.58

Base text production latency		0.66			0.43
Log odds of backspacing actions			0.90		0.72
Relative backspacing time			0.86		0.85
Keystroke efficiency	0.33		−0.37	−0.21	0.48
Number of fast bursts	−0.36			0.76	0.91
Total time at phrasal burst start	0.24			0.39	0.15
Total time between phrasal bursts				0.78	0.51

Note. Factor loadings < .2 are suppressed.

Table 7 Rotated Factor Loadings (Pattern Matrix) and Communalities for the Form 2 Knowledge Elicitation Item, Based on a Factor Analysis With Promax Rotation for 15 Process Log Features

Process Log feature	Production effort	Editing effort	Keyboarding effort	Overall deliberation time	Communalities
Relative word-initial time	1.02	0.21			0.89
Relative productive time	−0.92			−0.20	0.80
Relative start time	0.93				0.86
Word start speed	0.67		0.22		0.52
Maximum phrasal burst length	−0.40			0.25	0.38
Duration of within-word pauses	−0.24		0.81		0.63
Fast burst length			−0.60		0.40
Base text production latency			0.56		0.40
Duration of between-word pauses	0.49		0.62		0.72
Log odds of backspacing actions	0.22	0.90			0.72
Relative backspacing time		0.90			0.78
Keystroke efficiency	0.30	−0.51			0.56
Total time at phrasal burst start				0.77	0.51
Total time between phrasal bursts				0.91	0.74

Number of fast bursts	−0.46			0.52	0.75

Note. Factor loadings < .2 are suppressed.

Table 8 Rotated Factor Loadings (Pattern Matrix) and Communalities for the Form 3 Word Listing Item, Based on a Factor Analysis With Promax Rotation for 15 Process Log Features

Process Log feature	Production effort	Keyboarding effort	Editing effort	Overall deliberation time	Communalities
Relative word-initial time	.91				.84
Relative productive time	−.94	.43			.75
Relative start time	.88				.91
Word start speed	.80			.20	.68
Number of fast bursts	−.72			.38	.79
Duration of between-word pauses	.54	.27	−.24		.51
Maximum phrasal burst length	−.51				.35
Duration of within-word pauses		.75			.58
Fast burst length		−.71			.57
Base text production latency		.71			.53
Log odds of backspacing actions			.88		.72
Relative backspacing time			.83		.81
Keystroke efficiency	.35		−.47		.53
Total time at phrasal burst start	.29			.61	.25
Total time between phrasal bursts				.46	.40

Note. Factor loadings < .2 are suppressed.

Table 9 Rotated Factor Loadings (Pattern Matrix) and Communalities for the Form 3 Knowledge Elicitation Item, Based on a Factor Analysis With Promax Rotation for 15 Process Log Features

Process Log feature	Production effort	Keyboarding effort	Editing effort	Overall deliberation time	Communalities
Relative word-initial time	.74			.21	.85
Relative productive time	−.88	.21			.77
Relative start time	.98				.92
Word start speed	.46			.66	.83
Number of fast bursts	−.96			.29	.70
Duration of between-word pauses	.36	.37		.33	.76
Maximum phrasal burst length	−.53	−.30			.45
Duration of within-word pauses		.83			.62
Fast burst length		−.80			.58
Base text production latency		.54			.30
Log odds of backspacing actions			.74		.49
Relative backspacing time			.94		.95
Keystroke efficiency	.43		−.40		.48
Total time at phrasal burst start				.76	.46
Total time between phrasal bursts				.45	.29

Note. Factor loadings < .2 are suppressed.

Table 10 Rotated Factor Loadings (Pattern Matrix) and Communalities for the Form 4 Word Listing Item, Based on a Factor Analysis With Promax Rotation for 15 Process Log Features

Process Log feature	Production effort	Keyboarding effort	Editing effort	Overall deliberation time	Communalities
Relative word-initial time	.98		.20	.21	.81
Relative productive time	−.95	.28			.83

Relative start time	.84		.22	−.28	.80
Word start speed	.76				.64
Number of fast bursts	−.61			.44	.78
Duration of between-word pauses	.46	.41	−.22		.59
Maximum phrasal burst length	−.44	−.30			.35
Duration of within-word pauses		.72			.51
Fast burst length		−.76			.58
Base text production latency		.74			.51
Log odds of backspacing actions	.21		.93		.79
Relative backspacing time			.89		.78
Keystroke efficiency	.30	.20	−.53		.66
Total time at phrasal burst start	.23	.21		.43	.26
Total time between phrasal bursts				.66	.41

Note. Factor loadings < .2 are suppressed.

Table 11 Rotated Factor Loadings (Pattern Matrix) and Communalities for the Form 4 Knowledge Elicitation Item, Based on a Factor Analysis With Promax Rotation for 15 Process Log Features

Process Log feature	Production effort	Keyboarding effort	Editing effort	Overall deliberation time	Communalities
Relative word-initial time	0.94				0.79
Relative productive time	−1.02			−0.30	0.92
Relative start time	0.91				0.86
Word start speed	0.61	0.37			0.68
Number of fast bursts	−0.58	0.21		0.58	0.87
Maximum phrasal burst length	−0.47				0.33
Duration of between-word pauses	0.37				0.79
Duration of within-word pauses		0.76			0.55

Fast burst length	−0.76		−0.33	0.66
Base text production latency	0.61	0.23		0.50
Log odds of backspacing actions		0.76		0.60
Relative backspacing time		0.96		0.94
Keystroke efficiency	0.32	−0.38		0.47
Total time at phrasal burst start			0.38	0.15
Total time between phrasal bursts			0.55	0.30

Note. Factor loadings < .2 are suppressed.

For each of the eight items, the first factor can be interpreted as a measure of production effort. Positively loading features include relative start time, which can be viewed as a measure of effort generating ideas and text before typing begins, and relative word-initial time, which can be viewed as a measure of effort accessing words during text generation. Negatively loading features include relative productive time (the extent to which text production outpaces pauses to generate ideas), word start speed (rate of text production at word boundaries where lexical access might naturally slow down typing), the number of fast bursts (reflecting how consistently a writer is able to generate more text at any given point), maximum phrasal burst length (reflecting fluency of text production while generating a single phrase), and duration of between-word pauses (reflecting effort spent on generating the next word or phrase).

The second and third factors can be interpreted as measures of keyboarding effort and editing effort. For keyboarding effort, positively loading features include the base latency of all text production events and, specifically, the duration of in-word pauses, which reflect the vast majority of typing events not likely to be affected by higher order writing processes. Negatively loading features include the length of bursts of fast text production, which should reflect general fluency of transcription processes. For editing effort, positively loading features include the log odds of backspacing actions and the relative time spent on backspacing compared to text production. Negatively loading features include keystroke efficiency, which should be high if relatively little backspacing or other deletions occur.

Finally, the fourth factor can be interpreted as a measure of deliberation time. The most important features positively loading on this factor include the total time spent at the start of a phrasal burst or between phrasal bursts—in other words, time spent on the very longest pauses between bursts of text production.

Correlations Among Item Scores and Item Factors

After Promax rotation, correlations between the factors were generally weak (see Tables 12–15), though a few cross-factor correlations ranged as high as .55. In particular, the production effort and deliberation time variables had correlations of .55 in Form 1, knowledge elicitation; −.37 in Form 2, word listing; .50 in Form 3, knowledge elicitation; and .39 in Form 4, knowledge elicitation. A few forms had moderate correlations between the production effort and keyboarding effort factors, with correlations of .50 in Form 3, knowledge elicitation, and .43 in Form 4, knowledge elicitation.

Table 12 Correlations Among Item and Estimated Factor Scores for Form 1

Score	Word listing score	Revised Word listing score	Multiple-choice score	Production effort (WL)	Keyboarding effort (WL)	Editing effort (WL)	Deliberation Time (WL)	Production effort (KE)	Keyboarding effort (KE)	Editing effort (KE)	Deliberation time (KE)
Knowledge elicitation score	.39**	.42**	.32**	−.25**	−.14**	−.02	.17**.	−.48**	−.13*	.08	−.04
Word listing score		.89**	.29**	−.38**	−.09	−.07	.18**	−.15*	−.06	−.08	.05
Revised word listing score			.34**	−.31**	−.01	−.03	.19**	−.23**	−.06	.16*	.05
Multiple-choice score				−.19**	−.11*	−.05	.13*	−.10	−.02	.09	.11.
Production effort (WL)					.21**	−.17**	.11*	.23**	−.01	−.03	−.07
Keyboarding effort (WL)						−.09	.01	.04	.46**	−.07	.04
Editing effort (WL)							.30**	.07	−.04	.10	.10
Deliberation time (WL)								−.18**	.03	.14*	.02
Production effort (KE)									.25**	−.24**	.55**
Keyboarding effort (KE)										−.05	.33
Editing effort (KE)											.08

Note. KE = knowledge elicitation. WL = word listing.

*$p < .05$. **$p < .01$.

Table 13 Correlations Among Item and Estimated Factor Scores for Form 2

Score	Word listing score	Revised word listing score	Multiple-choice score	Production effort (WL)	Keyboarding effort (WL)	Editing effort (WL)	Deliberation Time (WL)	Production effort (KE)	Keyboarding effort (KE)	Editing effort (KE)	Deliberation time (KE)
Knowledge elicitation score	.53**	.56**	.42**	−.25**	−.18**	.04	.32**	−.45**	−.18**	.20**	.43**
Word listing score		.71**	.47**	−.48**	−.27**	.02	.59**	−.28**	−.25**	.08	.21**
Revised word listing score			.53**	−.34**	−.23**	−.02	.41**	−.36**	−.27**	.11	.24**
Multiple-choice score				−.25**	−.15**	−.07	.27**	−.31**	.16**	−.10	.16**
Production effort (WL)					.14**	−.19**	−.37**	.38**	.20**	.01	.01
Keyboarding effort (WL)						.02	−.21**	.13	.43**	−.00	−.01
Editing effort (WL)							.28**	−.15**	.14**	−.09	.02
Deliberation time (WL)								−.27**	.13*	.09	.24**
Production effort (KE)									−.31**	.20**	−.27**
Keyboarding effort (KE)										.18**	.44**
Editing effort (KE)											.12*

Note. KE = knowledge elicitation. WL = word listing.

*$p < .05$. **$p < .01$.

Table 14 Correlations Among Item and Estimated Factor Scores for Form 3

Score	Word listing score	Revised word listing score	Multiple-choice score	Production effort (WL)	Keyboarding effort (WL)	Editing effort (WL)	Deliberation Time (WL)	Production effort (KE)	Keyboarding effort (KE)	Editing effort (KE)	Deliberation time (KE)
Knowledge elicitation score	.47**	.49**	.35**	−.31**	−.19**	.02	.26**	−.45**	−.34**	.04	−.12*
Word listing score		.91**	.46**	−.54**	−.31**	−.08	.37**	−.35**	−.38**	.00	−.09
Revised word listing score			.54**	−.53**	−.28**	−.04	.38**	−.40**	−.39**	−.00	−.11*
Multiple-choice score				−.30**	−.11*	−.05	.30**	−.37**	−.20**	−.07	−.09
Production effort (WL)					.39**	−.17**	−.29**	.37**	.32**	−.03	.13*
Keyboarding effort (WL)						.30**	.10	.24**	.52**	.01	.27**
Editing effort (WL)							.34**	.01	.11*	.21**	.16**
Deliberation time (WL)								−.27**	.09	.19**	.13*
Production effort (KE)									.50**	−.29**	.47**
Keyboarding effort (KE)										.15**	.53**
Editing effort (KE)											.08

Note. KE = knowledge elicitation. WL = word listing.

*$p < .05$. **$p < .01$.

Table 15 Correlations Among Item and Estimated Factor Scores for Form 4

Score	Word listing score	Revised word listing score	Multiple-choice score	Production effort (WL)	Keyboarding effort (WL)	Editing effort (WL)	Deliberation Time (WL)	Production effort (KE)	Keyboarding effort (KE)	Editing effort (KE)	Deliberation time (KE)
Knowledge elicitation score	.58**	.53**	.30**	−.21**	−.07	.09	.32**	−.40**	−.16**	.11*	.32**
Word listing score		.71**	.31**	−.28**	−.07	−.04	.30**	−.28**	−.15**	−.02	.12*
Revised word listing score			.44**	−.24**	.02	.03	.26**	−.26**	−.10	.02	.16**
Multiple-choice score				−.13**	.02	−.01	.06	−.18**	−.14**	.02	.10*
Production effort (WL)					.31**	−.28**	−.28**	.32**	.09	.04	−.10
Keyboarding effort (WL)						−.10	−.10	.13*	.44**	−.05	.09
Editing effort (WL)							.34**	−.14*	−.10	.16**	.12*
Deliberation time (WL)								−.31**	−.01	.12*	.39**
Production effort (KE)									.43**	−.26**	−.39**
Keyboarding effort (KE)										.06	−.05
Editing effort (KE)											.28**

Note. KE = knowledge elicitation. WL = word listing.

*$p < .05$. **$p < .01$.

Across the two items on the same form, the production effort factors showed weak to moderate cross-correlations (.23 for Form 1, .38 for Form 2, .37 for Form 3, and .32 for Form 4). The keyboarding effort factors showed moderate correlations (.46 for Form 1, .43 for Form 2, .52 for Form 3, and .44 for Form 4). The editing effort factors were at most weakly correlated (nonsignificant on Forms 1 and 2, .21 on Form 3, and .16 on Form 4). The deliberation effort factors varied from no correlation to moderately correlated (nonsignificant on Form 1, .24 on Form 2, .13 on Form 4, and .39 on Form 4).

The correlations between scores for the word listing and knowledge elicitation items were moderate (.39 for Form 1, .53 for Form 2, .47 for Form 3, and .58 for Form 4.), suggesting they were measuring related, but not identical, constructs. There were also weak to moderate correlations between the multiple-choice scores and the constructed-response scores. The knowledge elicitation item scores were correlated with multiple-choice scores at .32 for Form 1, .42 for Form 2, .35 for Form 3, and .30 for Form 4. The word listing item scores were correlated with multiple-choice scores at .29 for Form 1, .47 for Form 2, .47 for Form 3, and .31 for Form 4, suggesting the tasks were measuring a related, but not identical, construct. The revised word listing items showed notable increases in the strength of correlation with multiple-choice scores, with correlations of .34 for Form 1 (vs. .29 unrevised), .53 for Form 2 (vs. .47 unrevised), .54 for Form 3 (vs. .46 unrevised), and .44 for Form 4 (vs. .31 unrevised). This suggests that students were either primed by the multiple-choice items or learned some of the terms by the time they answered the constructed responses.

In each form, there were moderate correlations between the knowledge elicitation production effort factor and scores for the knowledge elicitation item (−.48 for Form 1, .−.45 for Form 2, −.45 for Form 3, and −.40 for Form 4). Correlations between the word listing production effort factor and scores for the word listing items ranged between weak and moderate (−.38 for Form 1, −.48 for Form 2, −.54 for Form 3, and −.28 for Form 4).

The strength of the correlation between the deliberation time factor and scores for the knowledge elicitation and word listing items varied quite a bit across forms (for word listing, .18 for Form 1, .59 for Form 2, .37 for Form 3, and .30 for Form 4; for knowledge elicitation, nonsignificant for Form 1, .43 for Form 2, nonsignificant for Form 3, and .32 for Form 4).

The strength of the correlation between the keyboarding effort factor and scores for the corresponding item ranged from nonsignificant to weak (for word listing, nonsignificant on Form

1, −.27 on Form 2, −.31 on Form 3, and nonsignificant on Form 4; for knowledge elicitation, −.34 on Form 1, −.18 on Form 2, −.34 on Form 3, and −.16 on Form 4).

Finally, the editing effort factor showed little correlation with score. It had nonsignificant on all four forms for the word listing item. For the knowledge elicitation item, this correlation was nonsignificant on Form 1, .20 on Form 2, nonsignificant on Form 3, and .11 on Form 4.

Associations With Item Scores

We conducted a series of hierarchical multiple linear regressions to determine the influence of the process features on the knowledge scores. More specifically, we performed eight hierarchical multiple linear regressions in which we entered the estimated factor scores plus the multiple-choice score to predict the constructed responses. The variables were entered in the following order:

1. editing effort (other item); editing effort (same item)
2. keyboarding effort (other item); keyboarding effort (same item)
3. multiple-choice knowledge score
4. deliberation time (other item)
5. production effort (for the other item)
6. deliberation time (for the same item)
7. production effort (for the same item)

We reasoned that editing effort represents a baseline for the ability to monitor the quality of transcription (though this should not be particularly relevant to the score), that keyboarding effort represents a baseline for general transcription ability (only marginally relevant to the score), and that the multiple-choice score provides a baseline for general knowledge about the topic. The deliberation time and production effort factors for the opposite item provide a productive measure relevant to the topic, but not specific to the task, whereas the deliberation time and production effort factors for the same item measure the writer's ability to retrieve and output information about information that is specifically relevant to that item. Entering the items in this order will enable us to get a sense how much of the variance is accounted for by receptive knowledge of the topic and of topic-specific productive and task-specific factors. The details of these analyses are shown in Tables 16–24.

Table 16 Hierarchical Linear Regression Predicting Item Score for the Word Listing Item in Form 1

Factor score	*b*	*SE b*	Regression coefficient
Step 1			
Constant***	3.66	0.25	
Editing effort (other task; n.s.)	0.40	0.27	0.10
Editing effort (same task; n.s.)	−0.12	0.26	−0.03
Step 2			
Constant***	3.72	0.25	
Editing effort (other task; n.s.)	0.37	0.27	0.09
Editing effort (same task; n.s.)	−0.14	0.26	−0.04
Keyboarding effort (other task; n.s.)	−0.02	0.29	−0.00
Keyboarding effort (same task; n.s.)	−0.49	0.30	−0.12
Step 3			
Constant**	1.91	0.71	
Editing effort (other task; n.s.)	0.30	0.26	0.07
Editing effort (same task; n.s.)	−0.14	0.26	−0.03
Keyboarding effort (other task; n.s.)	−0.05	0.29	−0.01
Keyboarding effort (same task; n.s.)	−0.37	0.30	−0.09
Multiple-choice score**	0.15	0.06	0.17
Step 4			
Constant**	1.97	0.71	
Editing effort (other task; n.s.)	0.28	0.26	0.07
Editing effort (same task; n.s.)	−0.17	0.26	−0.04
Keyboarding effort (other task; n.s.)	−0.18	0.31	−0.05
Keyboarding effort (same task; n.s.)	−0.31	0.30	−0.07
Multiple-choice score**	0.15	0.06	0.17
Deliberation time (other task)	0.32	0.31	0.07
Step 5			
Constant**	2.27	0.71	
Editing effort (other task; n.s.)	0.03	0.28	0.01
Editing effort (same task; n.s.)	−0.24	0.25	−0.06
Keyboarding effort (other task; n.s.)	−0.15	0.31	−0.04
Keyboarding effort (same task; n.s.)	−0.31	0.30	−0.07
Multiple-choice score*	0.12	0.06	0.14
Deliberation time (other task)*	0.84	0.36	0.19

Production effort (other task)*	−0.87	0.32	−0.22
Step 6			
Constant**	2.42	0.71	
Editing effort (other task; n.s.)	0.00	0.27	0.00
Editing effort (same task; n.s.)	−0.42	0.26	−0.10
Keyboarding effort (other task; n.s.)	−0.18	0.31	−0.05
Keyboarding effort (same task; n.s.)	−0.34	0.30	−0.08
Multiple-choice score (n.s.)	0.11	0.06	0.12
Deliberation time (other task)*	0.80	0.35	0.18
Production effort (other task) *	−0.74	0.32	−0.19
Deliberation time (same task)*	0.68	0.27	0.17
Step 7			
Constant***	2.53	0.63	
Editing effort (other task; n.s.)	0.07	0.25	0.02
Editing effort (same task)***	−0.98	0.25	−0.24
Keyboarding effort (other task)*	−0.65	0.28	−0.16
Keyboarding effort (same task; n.s.)	0.26	0.28	0.06
Multiple-choice score*	0.10	0.05	0.12
Deliberation time (other task)*	0.77	0.32	0.18
Production effort (other task; n.s.)	−0.09	0.30	−0.02
Deliberation time (same task)***	1.42	0.26	0.34
Production effort (same task)***	−1.94	0.26	−0.50

Note. n.s. = not significant. R^2 = .00 for Step 1 (n.s.). ΔR^2 = .01 for Step 2 (n.s.), +.03 for Step 3 ($p < .05$), +.00 for Step 4 ($p < .05$), +.03 for Step 5 ($p < .01$), +.02 for Step 6 ($p < .01$), and +.18 for Step 7 ($p < .001$).
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 17 Hierarchical Linear Regression Predicting Item Score for the Knowledge Elicitation Item in Form 1

Factor score	*b*	*SE b*	Regression coefficient
Step 1			
Constant***	3.05	0.23	
Editing effort (other task; n.s.)	0.17	0.24	0.05
Editing effort (same task; n.s.)	0.30	0.25	0.08
Step 2			
Constant***	3.07	0.23	
Editing effort (other task; n.s.)	0.15	0.24	0.04
Editing effort (same task; n.s.)	0.28	0.25	0.07

Keyboarding effort (other task; n.s.)	−0.16	0.28	−0.04
Keyboarding effort (same task; n.s.)	−0.34	0.27	−0.09
Step 3			
Constant (n.s.)	0.90	0.65	
Editing effort (same task; n.s.)	0.15	0.23	0.04
Editing effort (other task; n.s.)	0.20	0.24	0.05
Keyboarding effort (other task; n.s.)	−0.02	0.27	−0.01
Keyboarding effort (same task; n.s.)	−0.38	0.26	−0.10
Multiple-choice score***	0.18	0.05	0.23
Step 4			
Constant (n.s.)	1.14	0.64	
Editing effort (other task; n.s.)	−0.07	0.24	−0.02
Editing effort (same task; n.s.)	0.12	0.24	0.03
Keyboarding effort (other task; n.s.)	−0.06	0.27	−0.02
Keyboarding effort (same task; n.s.)	−0.40	0.26	−0.11
Multiple-choice score**	0.16	0.05	0.20
Deliberation time (other task)***	0.81	0.25	0.21
Step 5			
Constant (n.s.)	1.22	0.63	
Editing effort (other task; n.s.)	−0.27	0.25	−0.07
Editing effort (same task; n.s.)	0.08	0.24	0.02
Keyboarding effort (other task; n.s.)	0.13	0.27	0.04
Keyboarding effort (same task)*	−0.51	0.26	−0.14
Multiple-choice score**	0.15	0.05	0.19
Deliberation time (other task)**	10.04	0.26	0.27
Production effort (other task; n.s.)	−0.72	0.24	−0.20
Step 6			
Constant**	1.25	0.64	
Editing effort (other task; n.s.)	−0.29	0.25	−0.08
Editing effort (same task; n.s.)	0.08	0.24	0.02
Keyboarding effort (other task; n.s.)	0.17	0.28	0.04
Keyboarding effort (same task)**	−0.57	0.28	−0.15
Multiple-choice score**	0.15	0.05	0.19
Deliberation time (other task)***	1.05	0.26	0.28
Production effort (other task)**	−0.74	0.25	−0.20
Deliberation time (same task; n.s.)	0.15	0.28	0.04
Step 7			

Constant**	1.85	0.56	
Editing effort (other task; n.s.)	−0.23	0.22	−0.06
Editing effort (same task)**	−0.53	0.22	−0.14
Keyboarding effort (other task; n.s.)	0.01	0.25	0.00
Keyboarding effort (same task; n.s.)	−0.33	0.25	−0.09
Multiple-choice score*	0.10	0.04	0.13
Deliberation time (other task)*	0.55	0.23	0.15
Production effort (other task; n.s.)	−0.18	0.23	−0.05
Deliberation time (same task)***	1.37	0.28	0.34
Production effort (same task)***	−2.27	0.26	−0.61

Note. n.s. = not significant. R^2 = .00 for Step 1 (n.s.). ΔR^2 = .01 for Step 2 (n.s.), +.05 for Step 3 ($p < .01$), +.03 for Step 4 ($p < .001$), +.02 for Step 5 ($p < .001$), .00 for Step 6 ($p < .001$), and +.21 for Step 7 ($p < .001$).
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 18 Hierarchical Linear Regression Predicting Item Score for the Word Listing Item in Form 2

Factor score	*b*	*SE b*	Regression coefficient
Step 1			
Constant***	8.34	0.40	
Editing effort (other task; n.s.)	0.70	0.43	0.09
Editing effort (same task; n.s.)	−0.30	0.45	−0.04
Step 2			
Constant***	8.32	0.38	
Editing effort (other task)*	0.99	0.42	0.13
Editing effort (same task; n.s.)	−0.45	0.43	−0.06
Keyboarding effort (other task)**	−1.56	0.47	−0.19
Keyboarding effort (same task)**	−1.49	0.50	−0.17
Step 3			
Constant (n.s.)	−0.79	10.13	
Editing effort (same task; n.s.)	0.51	0.38	0.07
Editing effort (other task; n.s.)	−0.03	0.40	−0.00
Keyboarding effort (other task)*	−1.00	0.44	−0.12
Keyboarding effort (same task)**	−1.27	0.45	−0.15
Multiple-choice score***	0.60	0.07	0.41
Step 4			
Constant (n.s.)	−0.41	10.13	
Editing effort (other task; n.s.)	0.01	0.42	0.00

Editing effort (same task; n.s.)	0.01	0.39	0.00
Keyboarding effort (other task)*	−1.09	0.43	−0.13
Keyboarding effort (same task)**	−1.24	0.45	−0.14
Multiple-choice score***	0.58	0.07	0.39
Deliberation time (other task)**	1.20	0.41	0.15
Step 5			
Constant (n.s.)	−0.01	10.16	
Editing effort (other task; n.s.)	−0.14	0.43	−0.02
Editing effort (same task; n.s.)	−0.07	0.39	−0.01
Keyboarding effort (other task)*	−0.95	0.44	−0.12
Keyboarding effort (same task)**	−1.24	0.45	−0.14
Multiple-choice score***	0.55	0.07	0.37
Deliberation time (other task)**	1.12	0.42	0.14
Production effort (other task; n.s.)	−0.57	0.40	−0.08
Step 6			
Constant*	2.13	10.01	
Editing effort (other task; n.s.)	0.13	0.37	0.02
Editing effort (same task)**	−1.18	0.35	−0.15
Keyboarding effort (other task)**	−1.11	0.38	−0.14
Keyboarding effort (same task; n.s.)	0.60	0.39	−0.07
Multiple-choice score***	0.40	0.06	0.27
Deliberation time (other task; n.s.)	0.38	0.36	0.05
Production effort (other task; n.s.)	−0.16	0.35	−0.02
Deliberation time (same task)***	3.84	0.35	0.49
Step 7			
Constant**	2.64	0.96	
Editing effort (other task; n.s.)	0.36	0.35	0.05
Editing effort (same task)***	−1.39	0.33	−0.17
Keyboarding effort (other task)***	−0.98	0.36	−0.12
Keyboarding effort (same task; n.s.)	−0.60	0.37	−0.07
Multiple-choice score***	0.36	0.06	0.24
Deliberation time (other task; n.s.)	0.65	0.35	0.08
Production effort (other task; n.s.)	0.51	0.34	0.07
Deliberation time (same task)***	3.28	0.34	0.42
Production effort (same task)***	−2.22	0.34	−0.28

Note. n.s. = not significant. $R^2 = .00$ for Step 1 (n.s.). $\Delta R^2 = +.09$ for Step 2 ($p < .001$), +.15 for Step 3 ($p < .001$), +.02 for Step 4 ($p < .001$), +.00 for Step 5 ($p < .001$), +.19 for Step 6 ($p < .001$), and +.06 for Step 7 ($p < .001$).
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 19 Hierarchical Linear Regression Predicting Item Score for the Knowledge Elicitation Item in Form 2

Factor score	*b*	*SE b*	Regression coefficient
Step 1			
Constant***	6.41	0.36	
Editing effort (other task; n.s.)	−0.24	0.40	−0.03
Editing effort (same task)***	1.49	0.39	0.21
Step 2			
Constant***	6.40	0.35	
Editing effort (other task; n.s.)	−0.38	0.40	−0.05
Editing effort (same task)***	1.73	0.38	0.24
Keyboarding effort (other task; n.s.)	−0.82	0.46	−0.10
Keyboarding effort (same task)***	−1.34	0.44	−0.18
Step 3			
Constant (n.s.)	1.57	1.05	
Editing effort (same task; n.s.)	−0.02	0.37	0.00
Editing effort (other task)***	1.32	0.36	0.18
Keyboarding effort (other task; n.s.)	−0.63	0.42	−0.08
Keyboarding effort (same task)**	−0.85	0.41	−0.12
Multiple-choice score***	0.53	0.07	0.39
Step 4			
Constant (n.s.)	−0.79	1.07	
Editing effort (other task; n.s.)	−0.35	0.38	−0.05
Editing effort (same task)***	1.27	0.35	0.18
Keyboarding effort (other task; n.s.)	−0.43	0.42	−0.05
Keyboarding effort (same task)*	−0.89	0.40	−0.12
Multiple-choice score***	0.47	0.07	0.35
Deliberation time (other task)**	1.12	0.37	0.16
Step 5			
Constant (n.s.)	−0.61	1.08	
Editing effort (other task; n.s.)	−0.41	0.38	−0.05
Editing effort (same task)***	1.30	0.35	0.18
Keyboarding effort (other task; n.s.)	−0.43	0.42	00.06
Keyboarding effort (same task)*	−0.83	0.40	−0.11
Multiple-choice score***	0.46	0.07	0.34

Deliberation time (other task)***	1.00	0.38	0.14
Production effort (other task; n.s.)	−0.45	0.38	−0.06
Step 6			
Constant (n.s.)	−0.05	1.01	
Editing effort (other task; n.s.)	−0.20	0.36	−0.03
Editing effort (same task; n.s.)	0.23	0.36	0.03
Keyboarding effort (other task; n.s.)	−0.47	0.39	−0.06
Keyboarding effort (same task)**	−0.96	0.38	−0.13
Multiple-choice score***	0.42	0.06	0.31
Deliberation time (other task; n.s.)	0.45	0.37	0.06
Production effort (other task; n.s.)	−0.65	0.35	−0.09
Deliberation time (same task)***	2.60	0.37	0.36
Step 7			
Constant (n.s.)	0.79	1.00	
Editing effort (other task; n.s.)	−0.30	0.35	−0.04
Editing effort (same task; n.s.)	−0.22	0.36	−0.03
Keyboarding effort (other task; n.s.)	−0.48	0.38	−0.06
Keyboarding effort (same task; n.s.)	−0.64	0.37	−0.09
Multiple-choice score***	0.37	0.06	0.27
Deliberation time (other task)***	0.42	0.36	0.06
Production effort (other task; n.s.)	−0.15	0.36	−0.02
Deliberation time (same task)***	2.35	0.36	0.32
Production effort (same task)***	−1.68	0.36	−0.24

Note. n.s. = not significant. $R^2 = .04$ for Step 1 ($p < .001$). $\Delta R^2 = +.05$ for Step 2 ($p < .001$), +.14 for Step 3 ($p < .001$), +.02 for Step 4 ($p < .001$), +.00 for Step 5 ($p < .001$), +.10 for Step 6 ($p < .001$), and +.04 for Step 7 ($p < .001$).
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 20 Hierarchical Linear Regression Predicting Item Score for the Word Listing Item in Form 3

Factor score	*b*	*SE b*	Regression coefficient
Step 1			
Constant***	10.79	0.76	
Editing effort (other task; n.s.)	0.21	0.80	0.01
Editing effort (same task; n.s.)	−1.34	0.83	−0.09
Step 2			
Constant***	10.49	0.71	

Editing effort (other task; n.s.)	0.71	0.76	0.05
Editing effort (same task; n.s.)	−0.21	0.81	−0.01
Keyboarding effort (other task)***	−4.32	0.94	−0.28
Keyboarding effort (same task)**	−2.73	0.96	−0.18
Step 3			
Constant (n.s.)	−2.68	1.71	
Editing effort (same task; n.s.)	0.06	0.70	0.00
Editing effort (other task; n.s.)	0.06	0.74	0.00
Keyboarding effort (other task)***	−3.32	0.86	−0.22
Keyboarding effort (same task)**	−0.258	0.88	−0.17
Multiple-choice score***	0.85	0.10	0.39
Step 4			
Constant (n.s.)	−2.56	1.70	
Editing effort (other task; n.s.)	0.06	0.69	0.00
Editing effort (same task; n.s.)	−0.15	0.74	−0.01
Keyboarding effort (other task)***	−4.34	0.956	−0.28
Keyboarding effort (same task)**	−2.51	0.87	−0.16
Multiple-choice score***	0.84	0.10	0.39
Deliberation time (other task)**	2.07	0.86	0.13
Step 5			
Constant (n.s.)	1.20	0.174	
Editing effort (other task; n.s.)	−1.07	0.73	−0.07
Editing effort (same task; n.s.)	−1.43	0.73	−0.10
Keyboarding effort (other task)*	−2.37	0.99	−0.15
Keyboarding effort (same task)***	−0.326	0.83	−0.21
Multiple-choice score***	0.60	0.11	0.28
Deliberation time (other task)**	2.82	0.90	0.17
Production effort (other task)**	−2.59	0.93	−0.18
Step 6			
Constant (n.s.)	1.20	0.174	
Editing effort (other task; n.s.)	−1.07	0.73	−0.07
Editing effort (same task; n.s.)	−1.43	0.73	−0.10
Keyboarding effort (other task)*	−2.37	0.99	−0.15
Keyboarding effort (same task)***	−3.26	0.83	−0.21
Multiple-choice score***	0.60	0.11	0.28
Deliberation time (other task; n.s.)	1.58	0.88	0.10
Production effort (other task; n.s.)	−1.58	0.91	−0.11

Deliberation time (same task)***	4.83	0.86	0.29
Step 7			
Constant (n.s.)	3.23	1.64	
Editing effort (other task; n.s.)	−0.25	0.69	−0.02
Editing effort (same task)***	−2.94	0.71	−0.20
Keyboarding effort (other task)**	−0.275	0.92	−0.18
Keyboarding effort (same task; n.s.)	−0.54	0.86	−0.04
Multiple-choice score***	0.47	0.10	0.22
Deliberation time (other task; n.s.)	1.47	0.82	0.09
Production effort (other task; n.s.)	−0.24	0.86	−0.02
Deliberation time (same task)***	3.86	0.81	0.24
Production effort (same task)***	−5.44	0.75	−0.38

Note. n.s. = not significant. R^2 = .00 for Step 1 (n.s.). ΔR^2 = +.15 for Step 2 ($p < .001$), +.14 for Step 3 ($p < .001$), +.01 for Step 4 ($p < .001$), +.01 for Step 5 ($p < .001$), +.06 for Step 6 ($p < .001$), and +.09 for Step 7 ($p < .001$). *$p < .05$. **$p < .01$. ***$p < .001$.

Table 21 Hierarchical Linear Regression Predicting Item Score for the Knowledge Elicitation Item in Form 3

Factor score	*b*	*SE b*	Regression coefficient
Step 1			
Constant***	11.47	0.73	
Editing effort (other task; n.s.)	−0.06	0.80	0.00
Editing effort (same task; n.s.)	0.57	0.77	0.04
Step 2			
Constant***	11.18	0.69	
Editing effort (other task; n.s.)	0.49	0.79	0.04
Editing effort (same task; n.s.)	1.20	0.75	0.09
Keyboarding effort (other task; n.s.)	−0.71	0.95	−0.05
Keyboarding effort (same task)***	−4.51	0.92	−0.31
Step 3			
Constant (n.s.)	2.35	1.77	
Editing effort (same task; n.s.)	0.68	0.76	0.05
Editing effort (other task; n.s.)	0.76	0.72	0.06
Keyboarding effort (other task; n.s.)	−0.60	0.91	−0.04
Keyboarding effort (same task)***	−3.84	0.89	−0.26
Multiple-choice score***	0.57	0.11	0.27
Step 4			

Constant*	4.06	1.82	
Editing effort (other task; n.s.)	−0.15	0.79	−0.01
Editing effort (same task)***	0.43	0.72	0.03
Keyboarding effort (other task; n.s.)	−0.96	0.90	−0.07
Keyboarding effort (same task)***	−3.35	0.89	−0.23
Multiple-choice score***	0.46	0.11	0.22
Deliberation time (other task)**	2.87	0.89	0.18
Step 5			
Constant (n.s.)	4.91	1.86	
Editing effort (other task; n.s.)	−0.60	0.82	−0.04
Editing effort (same task; n.s.)	0.54	0.72	0.04
Keyboarding effort (other task; n.s.)	−0.13	0.99	−0.01
Keyboarding effort (same task)***	−3.25	0.89	−0.22
Multiple-choice score***	0.40	0.11	0.20
Deliberation time (other task)**	2.51	0.91	0.16
Production effort (other task)*	−0.172	0.85	−0.13
Step 6			
Constant**	4.90	1.87	
Editing effort (other task; n.s.)	−0.62	0.82	−0.04
Editing effort (same task; n.s.)	0.55	0.72	0.04
Keyboarding effort (other task; n.s.)	−0.10	0.99	−0.01
Keyboarding effort (same task)**	−3.43	1.01	−0.23
Multiple-choice score***	0.40	0.11	0.20
Deliberation time (other task)**	2.43	0.93	0.16
Production effort (other task)*	−1.74	0.85	−0.13
Deliberation time (same task; n.s.)	0.34	0.90	0.02
Step 7			
Constant***	6.98	1.83	
Editing effort (other task; n.s.)	−0.09	0.80	−0.01
Editing effort (same task; n.s.)	−1.22	0.77	−0.09
Keyboarding effort (other task; n.s.)	−0.81	0.96	−0.06
Keyboarding effort (same task; n.s.)	−1.51	1.03	−0.10
Multiple-choice score*	0.27	0.11	0.13
Deliberation time (other task; n.s.)	1.68	0.90	0.11
Production effort (other task; n.s.)	0.79	0.83	−0.06
Deliberation time (same task)*	1.97	0.92	0.13
Production effort (same task)***	−5.19	0.97	−00.37

Note. n.s. = not significant. R^2 = .00 for Step 1 (n.s.). ΔR^2 = +.10 for Step 2 (p < .001), +.07 for Step 3 (p < .001), +.02 for Step 4 (p < .001), +.01 for Step 5 (p < .001), +.00 for Step 6 (p < .001), and +.06 for Step 7 (p < .001).
*p < .05. **p < .01. ***p < .001.

Table 22 Hierarchical Linear Regression Predicting Item Score for the Word Listing Elicitation Item in Form 4

Factor score	*b*	*SE b*	Regression coefficient
Step 1			
Constant***	7.83	0.38	
Editing effort (other task; n.s.)	0.16	0.39	0.02
Editing effort (same task; n.s.)	−0.31	0.41	−0.04
Step 2			
Constant***	7.80	0.38	
Editing effort (other task; n.s.)	0.12	0.39	0.02
Editing effort (same task; n.s.)	−0.37	0.41	−0.05
Keyboarding effort (other task; n.s.)	−0.33	0.59	−0.03
Keyboarding effort (same task; n.s.)	−0.42	0.47	−0.05
Step 3			
Constant (n.s.)	2.05	1.12	
Editing effort (same task; n.s.)	0.12	0.38	0.02
Editing effort (other task; n.s.)	−0.32	0.39	−0.04
Keyboarding effort (other task; n.s.)	0.24	0.57	0.03
Keyboarding effort (same task; n.s.)	−0.66	0.46	−0.08
Multiple-choice score***	0.38	0.07	0.29
Step 4			
Constant *	2.27	1.12	
Editing effort (other task; n.s.)	−0.12	0.41	−0.02
Editing effort (same task; n.s.)	−0.38	0.39	−0.05
Keyboarding effort (other task; n.s.)	0.06	0.59	0.01
Keyboarding effort (same task; n.s.)	−0.65	0.46	−0.08
Multiple-choice score***	0.36	0.07	0.27
Deliberation time (other task; n.s.)	0.72	0.46	0.09
Step 5			
Constant *	2.43	1.10	
Editing effort (other task; n.s.)	−0.43	0.39	−0.06
Editing effort (same task; n.s.)	−0.45	0.39	−0.06
Keyboarding effort (other task; n.s.)	0.68	0.60	0.07

Keyboarding effort (same task; n.s.)	−0.64	0.45	−0.08
Multiple-choice score***	0.35	0.07	0.26
Deliberation time (other task; n.s.)	0.04	0.49	0.01
Production effort (other task) ***	−1.86	0.49	−0.23
Step 6			
Constant *	2.42	1.06	
Editing effort (other task; n.s.)	−0.22	0.40	−0.03
Editing effort (same task) **	−1.03	0.39	−0.14
Keyboarding effort (other task; n.s.)	0.51	0.58	0.05
Keyboarding effort (same task; n.s.)	−0.38	0.43	−0.05
Multiple-choice score ***	0.35	0.07	0.26
Deliberation time (other task; n.s.)	−0.62	0.49	−0.08
Production effort (other task) **	−1.36	0.48	−0.17
Deliberation time (same task)***	2.33	0.46	0.29
Step 7			
Constant **	2.87	1.05	
Editing effort (other task; n.s.)	0.07	0.40	0.01
Editing effort (same task) **	−1.37	0.40	−0.18
Keyboarding effort (other task; n.s.)	0.20	0.57	0.02
Keyboarding effort (same task; n.s.)	0.10	0.45	0.01
Multiple-choice score***	0.32	0.07	0.24
Deliberation time (other task; n.s.)	−0.59	0.48	−0.07
Production effort (other task; n.s.)	−0.83	0.50	−0.10
Deliberation time (same task)***	2.17	0.45	0.27
Production effort (same task)***	−1.47	0.42	−0.20

Note. n.s. = not significant. R^2 = .00 for Step 1 (n.s.). ΔR^2 = +.00 for Step 2 (n.s.), +.07 for Step 3 ($p < .001$), +.00 for Step 4 ($p < .001$), +.04 for Step 5 ($p < .001$), +.06 for Step 6 ($p < .001$), and +.03 for Step 7 ($p < .001$).
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 23 Hierarchical Linear Regression Predicting Item Score for the Knowledge Elicitation Item in Form 4

Factor score	*b*	*SE b*	Regression coefficient
Step 1			
Constant***	8.27	0.35	
Editing effort (other task; n.s.)	0.44	0.38	0.06
Editing effort (same task)*	0.82	0.36	0.12
Step 2			

Constant***	8.21	0.35	
Editing effort (other task; n.s.)	0.37	0.38	0.05
Editing effort (same task)*	0.76	0.36	0.11
Keyboarding effort (other task; n.s.)	−0.26	0.44	−0.04
Keyboarding effort (same task; n.s.)	−0.73	0.54	−0.08
Step 3			
Constant**	3.05	1.03	
Editing effort (same task; n.s.)	0.41	0.36	0.06
Editing effort (other task)*	0.75	0.35	0.11
Keyboarding effort (other task; n.s.)	−0.47	0.42	−0.06
Keyboarding effort (same task; n.s.)	−0.21	0.53	−0.02
Multiple-choice score***	0.34	0.06	0.27
Step 4			
Constant**	3.34	0.99	
Editing effort (other task; n.s.)	−0.27	0.36	−0.04
Editing effort (same task; n.s.)	0.61	0.34	0.09
Keyboarding effort (other task; n.s.)	−0.21	0.41	00.03
Keyboarding effort (same task; n.s.)	−0.43	0.51	−0.05
Multiple-choice score***	0.32	0.06	0.26
Deliberation time (other task)***	2.34	0.39	0.31
Step 5			
Constant***	3.58	0.99	
Editing effort (other task; n.s.)	−0.42	0.37	−0.06
Editing effort (same task)*	0.68	0.34	0.10
Keyboarding effort (other task; n.s.)	0.01	0.42	0.00
Keyboarding effort (same task; n.s.)	−0.50	0.51	−0.06
Multiple-choice score***	0.30	0.06	0.24
Deliberation time (other task)***	2.20	0.40	0.29
Production effort (other task; n.s.)	−0.67	0.37	−0.10
Step 6			
Constant***	3.97	0.98	
Editing effort (other task; n.s.)	−0.37	0.37	−0.05
Editing effort (same task; n.s.)	0.19	0.36	0.02
Keyboarding effort (other task; n.s.)	−0.07	0.41	−0.01
Keyboarding effort (same task; n.s.)	−0.86	0.51	−0.09
Multiple-choice score***	0.27	0.06	0.22
Deliberation time (other task)***	1.68	0.42	0.22

Production effort (other task; n.s.)	−0.57	0.37	−0.08
Deliberation time (same task)***	1.61	0.43	0.21
Step 7			
Constant***	3.97	0.95	
Editing effort (other task; n.s.)	−0.23	0.36	00.03
Editing effort (same task; n.s.)	−0.24	0.36	−0.04
Keyboarding effort (other task; n.s.)	−0.26	0.41	−0.04
Keyboarding effort (same task; n.s.)	−0.16	0.52	−0.02
Multiple-choice score***	0.27	0.06	0.22
Deliberation time (other task)**	1.38	0.41	0.18
Production effort (other task; n.s.)	−0.07	0.38	−0.01
Deliberation time (same task)*	1.07	0.44	0.14
Production effort (same task)***	−1.98	0.45	−0.26

Note. n.s. = not significant. R^2 = .02 for Step 1 ($p < .05$). ΔR^2 = +.01 for Step 2 ($p < .05$), +.07 for Step 3 ($p < .001$), +.10 for Step 4 ($p < .001$), +.01 for Step 5 ($p < .001$), +.03 for Step 6 ($p < .001$), and +.04 for Step 7 ($p < .001$).
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 24 Paired-Sample *t*-Test: Mean Differences Between Word List and Knowledge Elicitation Feature Values

Factor and feature	Form 1		Form 2		Form 3		Form 4	
	Diff *M*	*SD*	Diff *M*	*SD*	Diff *M*	*SD*	Diff *M*	*SD*
Production effort								
Relative word-initial time	0.42*	1.26	0.26*	1.38	0.63*	1.49	0.57*	1.32
Relative productive time	−1.07*	1.77	−0.27*	1.24	−0.55*	1.29	−0.61*	1.09
Relative start time	0.30*	1.73	−0.21	1.87	0.87*	1.88	0.79*	1.58
Word start speed	5.10*	17.80	2.87*	11.03	6.61*	19.05	6.74*	16.21
Number of fast bursts	−1.44*	1.87	0.15	1.95	−1.62*	1.82	−1.90*	1.80
Duration of between-word pauses	0.81*	1.31	0.48*	1.02	0.92*	1.26	0.98*	1.21
Maximum phrasal burst length	−5.37*	5.75	−2.34*	5.76	−5.00*	4.95	−6.99*	6.14
Keyboarding effort								
Duration of within-word pauses	0.12*	0.31	0.11*	0.34	0.12*	0.29	0.16*	0.29
Fast burst length	−0.29*	0.52	−0.31*	0.60	−0.27*	0.51	−0.35*	0.50
Base text production latency	0.13*	0.56	0.02	1.25	0.00	1.50	0.17*	0.43
Editing effort								
Log odds of backspacing actions	−0.23	2.38	−0.08	1.05	0.07	1.03	0.17*	1.02
Relative backspacing time	0.06	1.05	0.02	2.61	−0.19	2.47	−0.12	2.42
Keystroke efficiency	−0.06	1.23	−0.14	1.39	0.26*	1.25	0.11	0.90
Deliberation time								
Total time at phrasal burst start	−1.13	34.10	14.01*	41.94	−8.86*	35.90	−4.46	31.32
Total time between phrasal bursts	8.15	41.46	8.33*	33.43	2.80	42.59	3.42	30.96

*$p < .003$ (significant for 16 comparisons after Bonferroni correction).

Overall, the editing effort factors did not make a significant initial contribution to R^2 (although the knowledge elicitation editing effort factor was a significant predictor in Step 1 for the knowledge elicitation items in Forms 2 and 4 and the same-task editing effort factor was a significant predictor in the final step of five of the eight regressions).

The keyboarding effort factors did not make a consistent contribution of additional variance to R^2 in all four forms, although they did make significant contributions ranging between an increase of +.05 and +.15 in R^2 in Forms 2 and 3 (both factors in the word listing items, with the knowledge elicitation keyboarding effort factor remaining significant in the final model, and the same-item keyboarding factor, in the knowledge elicitation items, where neither keyboarding effort factor remained significant in the final model).

The total multiple-choice score variable is a consistently significant contributor of additional variance to R^2. In the word listing items, it contributes +.03 in Form 1, +.15 in Form 2, +.14 in Form 3, and +.07 in Form 4. In the knowledge elicitation items, it contributes +.05 in Form 1, +.14 in Form 2, +.07 in Form 3, and +.07 in Form 4. The final regression coefficients are consistently significant. In the word listing items, multiple-choice score has a final regression coefficient of .12 in Form 1, .24 in Form 2, .22 in Form 3, and .24 in Form 4. In the knowledge elicitation items, multiple-choice score has a final regression coefficient of .13 in Form 1, .27 in Form 2, .13 in Form 3, and .22 in Form 4.

The deliberation time factor for the other item was a consistently significant, though small, contributor of additional variance to R^2. In the word listing items, it contributes less than +.01 in Form 1, +.02 in Form 2, +01 in Form 3, and less than +.01 in Form 4. In the knowledge elicitation items, it contributes +.03 in Form 1, +.02 in Form 2, +.02 in Form 3, and +.10 in Form 4. This factor's regression coefficients in the final model were small and not always significant. In the word listing items, the weights were .18 in Form 1, .08 in Form 2, .09 in Form 3, and nonsignificant in Form 4. In the knowledge elicitation items, the weights were .15 in Form 1, .06 in Form 2, ,11 in Form 3, and .18 in Form 4.

The production effort factor for the other item was a consistently significant, though small, contributor of additional variance to R^2. In the word listing items, it contributed +.03 in Form 1, less than +.01 to Form 2, +.01 to Form 3, and +.04 in Form 4. In the knowledge elicitation items, it contributed +.02 in Form 1, less than +.01 in Form 2, and +.01 in Forms 3 and 4. However, it was never a significant predictor in the final model, suggesting that it accounted for no variance distinct from that accounted for by the same-item deliberation time and production effort factors.

The deliberation time factor for the same item was a consistently significant contributor of additional variance to R^2. In the word listing items, it contributed +.02 in Form 1, +.19 in Form 2, and +.06 in Forms 3 and 4. In the knowledge elicitation items, it contributed less than +.01 in Form 1, +.10 in Form 2, less than +.01 in Form 3, and +.03 in Form 4. This factor also had significant weights in all eight final models. In the word listing items, the final regression coefficients were .34 in Form 1, .42 in Form 2, .24 in Form 3, and .27 in Form 4. In the knowledge elicitation items, the final regression coefficients were .34 in Form 1, .32 in Form 2, .13 in Form 3, and .14 in Form 4.

Finally, the production effort factor for the same item was a consistently significant contributor of additional variance to R^2. In the word listing items, it contributed +.18 in Form 1, +.06 in Form 2, +.09 in Form 3, and +.03 in Form 4. In the knowledge elicitation items, it contributed +.21 in Form 1, +.04 in Form 2, +.06 in Form 3, and +.04 in Form 4. This factor generally had the largest regression coefficient in the final model. In the word listing items, the final regression coefficients were −.50 in Form 1, −.28 in Form 2, −.38 in Form 3, and −.20 in Form 4. In the knowledge elicitation items, the final regression coefficients were −.61 in Form 1, −.24 in Form 2, −.37 in Form 3, and −.26 in Form 4.

Comparison of Means

A paired-samples t-test was conducted to compare the 16 features used in the factor analysis, with a Bonferroni correction of $.05/16 = p < .003$ to account for multiple comparisons. We found that there were several consistent differences between the two item types. In particular, (a) there were consistent differences between the two item types in the features associated with the production effort factor, with word listing items showing significantly longer pauses associated with production effort and significantly shorter bursts of fast text production than knowledge elicitation items, and (b) there were consistent differences between the two item types in the features associated with the keyboarding effort factor, with word listing items showing significantly longer within-word pauses, shorter bursts, and a slower overall production rate. There did not appear to be consistent, significant differences between the two item types on the editing effort and deliberation time features.

Discussion

Overall, the general pattern of results appeared to be consistent with what we would expect given the items' design.

The failure of the keyboarding effort and editing effort factors to contribute significantly to predicting scores is consistent with the intended (prior knowledge) construct, because the word listing and knowledge elicitation item types are supposed to measure background knowledge, not general writing fluency or writing quality.

The fact that the total multiple-choice score contributes significantly to predicting score in both item types is consistent with the assumption that all items within a form are testing for knowledge about the topic tested by that form. The fact that the other- or alternative-prompt production effort and deliberation time factors contribute significantly to score prediction is consistent with the assumption that the items require test takers to retrieve relevant information about the topic and express it productively. Finally, the fact that the same-prompt production effort and deliberation time factors makes sense on the assumption that the specific recall strategies employed in each task are relatively distinct, based on both the content to be provided and the specific response format used.

In addition, our findings are consistent with the assumption that students are answering word listing and knowledge elicitation items by following a knowledge-telling strategy, that is, by retrieving knowledge from memory and then expressing it sequentially, without putting significant effort into revising or restructuring that information once it has been written down.

If we assume that students are following a knowledge-telling strategy, we can account for nearly all of the facts we have adduced. Use of a knowledge-telling strategy explains why students spent so much time adding sequentially to their existing responses and spent little time cutting, pasting, or replacing existing text. It explains why the production effort and deliberation time factors account for the bulk of the variance covered by our multiple regression models. Finally, it explains why the production effort factors are negatively (and the deliberation time factors are positively) associated with score. Pauses between bursts can be interpreted as episodes during which information is retrieved from long-term memory and encoded for production during the next burst.

If we compare these results to the results of our prior studies of middle school essay writing (i.e., Deane, 2014; Deane & Zhang, 2015; Zhang & Deane, 2015), there appear to be strong similarities in the feature patterns associated with high and low performance. In both the present and in prior studies, the bulk of the variance in student scores can be accounted for using measures of planning and productivity (including total time, burst lengths, time spent before typing begins, and pauses between words). In both the present and prior studies, we identified keyboarding fluency and local editing factors, which accounted for much less of the total variance. These results are consistent with the hypothesis that in both populations, the typical writer deploys a knowledge-telling strategy.

One of the more striking results we observed was the presence of sharp differences in the sizes of the correlations between multiple-choice scores and word listing item scores on Forms 1 and 4 versus Forms 2 and 3 (in which the correlations were more than .15 higher). We hypothesize that this difference may be due to differences in the ways the questions were written. The multiple-choice forms included questions designed to measure all of the knowledge tested in the word listing and knowledge elicitation items. For Forms 2 and 3, the word listing and knowledge elicitation items measure very similar constructs. In particular, in Form 2, the knowledge indicated by the specific topic (the Emancipation Proclamation) and the general topic (the American Civil War) overlap heavily. Similarly, in Form 3, the general topic (colonial America) and the specific topic (the 13 original American colonies) are very closely related. By contrast, Form 1 requires the respondent to link the general topic, American immigration, with a topic that does not seem so closely related (diversity in America in the 19th century). Similarly, Form 4 requires the respondent to link the general topic (women's right to vote) with very specific knowledge about the history of feminism and the 19th Amendment. It thus seems reasonable to suppose that the word listing items are more likely to elicit vocabulary related to all of the multiple questions in Forms 2 and 3 (where the topics are more closely related) than in Forms 1 and 4, where there is more divergence between the focus of the two item types.

Finally, we observed significant differences in the values of writing process features that may correspond to differences in task requirements. Output format—whether or not one is producing lists of words or grammatically structured text—appeared to affect the range of values observed and, in some cases, the interpretation of a wide range of keystroke features. In particular, students appeared to be generally more fluent on the knowledge elicitation task, in which they were prompted to produce coherent text rather than lists of words. The task of producing a list of words not only appeared to produce more pauses at the beginning of the text and between words but also to slow down even the typing of individual words. By contrast, a more fluent performance seems to be evoked by the task of producing connected text expressing what one knows about a topic.

Limitations and Next Steps

The current study only examined data for four topics, all of them drawn from a single domain (history) and a single age range (high school). We did not directly compare student performance on these tasks with student performance on essay-writing tasks or other tasks (such as retyping someone else's text or editing/revising an existing document). It is therefore important to recognize that this study does not establish how student process data may shift when task conditions and expectations change, though it does establish an important baseline result when results from prior studies are taken into account—a common pattern of knowledge-telling behavior shared between exemplary short-answer and single-session, single-draft essay-writing tasks.

Another limitation of the current study was that the knowledge elicitation items were only presented in one position in each form: before all of the multiple-choice items. Given the increase in performance on the revised word listing items presented at the end of the form, it is quite possible that students would have demonstrated stronger performance on the knowledge elicitation items if they had attempted that task after the other items had been completed. That increase in performance might have provided a better estimate of what students actually know, because it would have had the effect of priming their knowledge of the topic and minimizing the effort needed to retrieve information the participants already knew but might have had difficulty retrieving (though such an effect might also reflect learning that takes place during the test session). We hope to examine this question in future studies.

Overall, a critical issue in keystroke log analysis of writing is the need to develop a much clearer account of how writing process features may shift in their distributions and require different interpretations in response to changes in task conditions and expectations. This study provided a first step toward establishing some baselines. We hope in future studies to examine how writing behavior changes across a variety of writing tasks.

Acknowledgments

The research reported here was supported by the U.S. Department of Education Institute of Education Sciences, Award No. R305A150176 to Educational Testing Service. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education or Educational Testing Service.

References

Almond, R., Deane, P., Quinlan, T., & Wagner, M. (2012). *A preliminary analysis of keystroke log data from a timed writing task* (Research Report No. RR-12-23). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2012.tb02305.x

Alves, R. A., Branco, M., Castro, S. L., & Olive, T. (2012). Effects of handwriting skill, output modes, and gender on fourth graders' pauses, language bursts, fluency, and quality. In V. Berninger (Ed.), *Past, present, and future contributions of cognitive writing research to cognitive psychology* (pp. 389–402). East Sussex, England: Psychology Press.

Alves, R. A., & Limpo, T. (2015). Progress in written language bursts, pauses, transcription, and written composition across schooling. *Scientific Studies of Reading, 19*, 374–391. https://doi.org/10.1080/10888438.2015.1059838

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hinsdale, NJ: Lawrence Erlbaum.

Berninger, V. W., Winn, W., MacArthur, C. A., Graham, S., & Fitzgerald, J. (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. In *Handbook of writing research* (pp. 96–114). New York, NY: Guilford Press.

Breetvelt, I., Van den Bergh, H., & Rijlaarsdam, G. (1994). Relations between writing processes and text quality: When and how? *Cognition and Instruction, 12*, 103–123. Retrieved from http://www.jstor.org/stable/3233677

Child, D. (2006). *The essentials of factor analysis*. New York, NY: Bloomsbury.

Connelly, V., Dockrell, J. E., Walter, K., & Critten, S. (2012). Predicting the quality of composition and written language bursts from oral language, spelling, and handwriting skills in children with and without specific language impairment. *Written Communication, 29*, 278–302. https://doi.org/10.1177/0741088312451109

Deane, P. (2014). *Using writing process and product features to assess writing quality and explore how those features relate to other literacy tasks* (Research Report No. RR-14-03). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12002

Deane, P., & Zhang, M. (2015). *Exploring the feasibility of using writing process features to assess text production skills* (Research Report No. RR-15-26). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12071

Graham, S., & Harris, K. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist, 35*, 3–12. https://doi.org/10.1207/S15326985EP3501_2

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication, 29*, 369–388. https://doi.org/10.1177/0741088312451260

Kaufer, D. S., Hayes, J. R., & Flower, L. (1986). Composing written sentences. *Research in the Teaching of English, 20*, 121–140.

Kellogg, R. T. (2001). Competition for working memory among writing processes. *American Journal of Psychology, 114*, 175–191. https://doi.org/10.2307/1423513

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review, 8*, 299–325. https://doi.org/10.1007/BF01464076

Miller, K. S. (2000). Academic writers on-line: Investigating pausing in the production of text. *Language Teaching Research, 4*, 123–148. https://doi.org/10.1177/136216880000400203

National Center for Education Statistics. (2014). *National assessment of educational progress (NAEP), 2014 science assessment*. Washington, DC: U.S. Department of Education.

Olive, T., Kellogg, R. T., & Piolat, A. (2008). Verbal, visual, and special working memory demands during text composition. *Applied Psycholinguistics, 29*, 669–687. https://doi.org/10.1017/S0142716408080284

O'Reilly T., Weeks, J., Sabatini, J., Halderman, L., & Steinberg, J. (2014). Designing Reading Comprehension Assessments for Reading Interventions: How a Theoretically Motivated Assessment Can Serve as an Outcome Measure. *Educational Psychology Review,* 26, (3), 403-424.

Snow, E. L., Allen, L. K., Jacovina, M. E., Perret, C. A., & McNamara, D. S. (2015). You've got style: Detecting writing flexibility across time. In *Proceedings of the fifth International*

Conference on Learning Analytics and Knowledge (pp. 194–202). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/2723576.2723592

Van Waes, L., Leijten, M., Wengelin, A., & Lindgren, E. (2011). Logging tools to study digital writing processes. In *Past, present, and future contributions of cognitive writing research to cognitive psychology* (pp. XX–XX). London, England: Psychology Press.

Wengelin, Å. (2006). Examining pauses in writing: Theory, methods and empirical data. In *Studies in writing: Vol. 18. Computer key-stroke logging and writing: Methods and applications* (pp. 107–130). New York, NY: Elsevier.

Zhang, M., & Deane, P. (2015). *Process features in writing: Internal structure and incremental value over product features* (Research Report No. RR-15-27). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12075

Suggested citation:

Deane, P., O'Reilly, T., Chao, S.-F., & Dreier, K. (2018). *Writing processes in short written responses to questions probing prior knowledge* (Research Report No. RR-18-xx). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.xxxx

Action Editor: James Carlson

Reviewers: Jesse Sparks and Jiangang Hao

[1] GISA is an acronym, for "global, integrated, scenario-based assessments," used to name a set of assessments of reading comprehension developed for the Reading for Understanding research initiative funded by the Institute for Education Sciences.