**Title:**

**Exploration of a Bayesian updating methodology on the data in a newly**

**developed computer-based assessment**

**Abstract**

**(---write about 170 words in this section---)**

**1. Introduction**

In educational research, discussions are often centered around writing instruction, writing competency, and writing assessment. The traditional (pencil–paper based) assessment adopts a single score provided by a human rater (e.g., a teacher) to measure a student’s writing competency. Human score is a holistic score. In a computer-based writing assessment, other than the holistic score, automated essay scoring (AES) systems can provide analytical scores to assess students’ text production skills. Generally speaking, these analytical scores are (1) features concerning grammatical and spelling errors (e.g., error counts normalized by the essay length), and (2) features based on discourse length and syntactic variety. The AES systems essentially rely on natural language processing (NLP), along with recently developed machines learning techniques (see a review paper published by Strobl et al., 2019). A set of relatively mature AES engines, for example, e-rater, EssayMetrics, have been integrated to writing instruction and learning (e.g., McNamara 2022).

As mentioned above, the human scores along with the automatic feature extraction approach help researchers and teachers understand students’ text production skills. These methods evaluate the essay quality based on submitted essays, which in a sense are *writing product* measures. In addition, in order to seek the mechanisms advancing writing development, a body of research focus on developing techniques to measure *writing process*. The quantification of writing process may contribute to establish a writing profile for each student. Such profiles consist of information collected from both the product and process of writing, together representing a student’s learning status (Xu & Qi, 2017), and will contain richer information to further influence the ongoing or future writing instruction (Baaijen & Galbraith, 2018).

**1.1. Related Work**

Within this scope, a log file in a computer platform stores time-stamped action sequences to facilitate later analysis. Keystroke log data can capture students’ writing behavior as they respond to writing assignment prompts (Chukharev-Hudilainen, 2014). A primary challenge in analyzing keystroke log data is centered around extracting meaningful features or variables. Most of the relevant studies have employed descriptive statistics, data reduction techniques, and machines learning methods. For example, pause frequencies during writing at different text boundaries (e.g., between sentences or between words) across writing tasks are meaningful (Medimorec & Risko, 2017; Conijn et al., 2019). Employing principal component analysis, the underlying dimensionality of the process data has been identified (Baaijen et al., 2012; Zhang & Deane, 2015). Uto et al. (2020) used a time- and learner-dependent hidden Markov model to analyze the writing process. Li (2020) used a mixture of lognormal models, the estimated parameters on pause data are consistent across two writing genres. The mixture model captures aspects of the writing process not examined otherwise. Namely, students with low human scores have a wide range of values on the mixing proportion parameter, whereas students with higher scores do not possess this pattern. Sinharay et al. (2019) used a regression tree to identify the association between writing process features and writing product features. Using a set of process variables, Sinharay et al. found that their machine learning technique slightly outperformed linear regression in predicting essay scores. In line with the current research trends in log data analysis, they pointed out the need of a variety of methods to “help explain the complex writing processes and validate the variables extracted from the writing processes for educational purpose” (p. 134).

**1. 2. Research Goal**

In this research, we intend to enrich the log data analysis literature by using a fully Bayesian approach. After examining the literature, we noted that distribution-based analysis is lacking. The advantage of researching the distributions of response time (e.g., pause data) is notable; that is, the parametric methods can provide meaningful summary statistics to represent each student. A fully Bayesian approach often refers to Bayesian updating methodology where the advantage, borrowing strength, can be used in hierarchical models. Using a hierarchical model involves placing a usually noninformative prior at the highest level of the hierarchy. Also, the hierarchical modeling, via the exchangeability assumption (i.e., the essays have a common prior), allows the essays “borrow strength” from one another, which in turn gain the precision (smaller variance).

The goal for this research is to extract meaningful summary statistics to represent each student’s writing process. Particularly, we will (1) build a computer-based assessment to capture students’ writing process, (2) classify the raw data into pause events, (3) illustrate how a Bayesian hierarchical mixture model can provide the interpretable statistics for the writing process.

Furthermore, we add value to the extant literature by researching the probability distribution of log data as well as highlighting the cognitive basis of process data. We endorse the idea that the cognitive models should be the foundation when analyzing log data; any identified features or unique patterns can thus be tied back to the cognitive model, for the sake of interpretability. Cognitive science and time data have been examined together through distribution analysis, probably because distribution-based measures can help determine the characteristics of indirect evidence about latent processes. Alternatively, the Gaussian distribution is less useful to model human response time. White and Staub (2012) studied the distribution of fixation durations during reading using an ex- Gaussian distribution. Zhang et al. (2018) explored response delay in terms of attention, by comparing individuals on the parameters extracted from the Gamma distribution. Palmer et al. (2011) employed a set of distributions including ex-Wald and Weibull distributions to interpret the mechanism in visual search. In the brief review below, we describe the cognitive model of writing and the related empirical evidence.

**1.3. Cognitive Model of Writing**

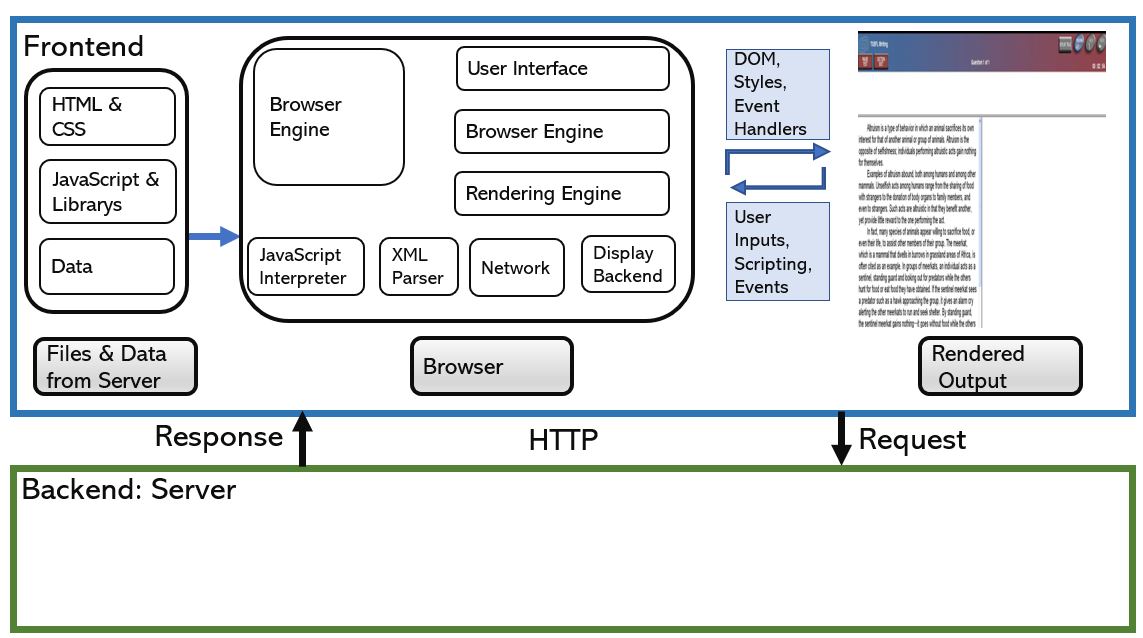
**(---write about 500 words in this section---)**

The rest of the article is organized as follows. In Section 2, we describe the design process of a computer-based writing assessment as well as how keystroke log data have been parsed into pause event. In Section 3, we propose the model to estimate the parameter for each student (data vector) in the context of Bayesian hierarchical model.  
Section 3 describes the estimated parameters based on MCMC algorithm. In Section 4, we interpret the results with the connection of the existing literature. In Section 5, we provide the conclusion of this research.

**2. The Present Study**

**2. 1. The development of a computer-based writing assessment**

**(---Junjie write about 400 words in this section---,basically, you only need to complete this Figure and describe this figure in words.)**

****

*Figure 1. The design of the assessment*

**2.2. Participants and Data Collection**

**(---write about 200 words in this section---)**

**2. 2. Pause Event Classification**

It is a computer-administered assessment. This means that it is relatively simple tocapture additional information about timing of data entry as part of the assessment. However, theraw stream of information is very large, and it needs to be processed in various ways. Inparticular, we want to be able to classify the various events in the log according to what kind ofprocessing the student was doing at the time. To this end, we want to classify each entry in thetiming log into one of six possible states:The event classification works by assuming that InWord*;* (Between) Word, Sentence,  
Paragraph, BackSpace, and Edit are states that the writer could be in. The classifier is then a  
finite state machine that works through the event stream classifying the events according to what state it is currently in.

**(Junjie keeps writing about 300 words in this section.)**

**2.3. The Proposed Model**

**(---write about 300 words in this section---)**

**2.4. Random-walk Metropolis Algorithm**

**(---write about 500 words in this section---)**

**3. Results**

**3.1 MCMC Convergence Diagnostics**

**(---write about 1000 words in this section---)**

**3.2 Model Selection**

**(---write about 500 words in this section---)**

**4. Discussion**

**(---write about 400 words in this section---)**

**5. Conclusion**

Although conducted with relatively small groups of students, these data sets allow us to define potential summary variables and algorithms for extracting them from keystroke logs and to test the robustness of those algorithms. We can also see which of the variables have meaningful amounts of variation and, hence, might be candidates for use in large-scale studies.

**(---write about 300 words in this section---)**

**Reference:**

**1000 words**