# Incorporating Pauses in Process Data Modeling with Heterogeneous Hidden Markov Models

Caitlin Tenison[1] (ctenison@ets.org) and Burcu Arslan[2]

[1]Educational Testing Service, Princeton, NJ, USA

[2]Educational Testing Service Global B.V., Amsterdam, Netherlands

## 1. Abstract
In the current study, we present heterogenous hidden Markov models as a promising method for modeling both the context and timing of responses. This work has implications for how incorporate response time into process data models of complex interactive computer tasks and reflect on the cognitive processes driving student behavior.

## 2. Introduction
The relationship between proficiency and length of pauses depends on the task and construct being assessed. For example, long pauses can reflect periods of planning [1] or wheel-spinning [2] and short pauses can indicate rapid guessing behavior [3,4] or a student is efficiently executing a well formulated plan. Prior work incorporating response time into the analysis of interactive computer tasks (ICTs) primarily analyzes the response-times of simple item-types [5]. Bergner and Von Davier [6], note that for complex tasks where many strategies exist to complete the task, the context of the actions taken is a critical dimension for using response time.

Although several studies have explored the use of hidden Markov models (HMMs) to model student's decision making in context [7,8], timing of actions and the pauses between them have not been incorporated into these models. In this study, we present a method for modeling pauses and actions in ICTs so that both choice timing and context is represented in these models.

Heterogeneous HMMs capture the probabilistic transition between latent states in sequential timesteps, applying the Markovian assumption that the current state is driven by the previous state and currently observed data (Figure 2C). Unlike discrete and gaussian HMMs, which only use a single type of data, HHMMs can use different distributions to estimate the emission probabilities of observed data (see [9] for further explanation). We explore how fitting HHMM to the student's action sequences can allow us to capture the context of actions by modeling the probabilistic transition between latent states and account for differences in the time it takes to produce those actions; two important indications of the cognitive processes underlying those actions.

## 3. Methods
### 3.1 Participants
We investigated data from student's interactions on the first question of a science ICT designed to assess their science inquiry practices related to the understanding of the concept of saturation (i.e., maximum concentration) and control-of-variables strategy. 164 sixth-grade ($N_{female} = 72$) and 131 eight-grade students' ($N_{female} = 67$) process data was captured while they worked in an interactive environment where they had the tools to run experiments, organize data and report conclusions (Figure 1).

### 3.2 HHMMs
For this task, we estimate the latent problem-solving states by fitting our HHMMs to three streams of observable data: Action events, proceeding pauses, and actor labels (Figure 2). For our action states, we translated task log files into a series of discrete timestamped actions that reflected student's inquiry process (Figure 2A). We chose action types that align with the top-level goal structure of prior cognitive models built to solve a similar task [10]. We

considered the pauses proceeding the actions as an indirect indication of the cognitive processing necessary to produce that action (Figure 2B). Since the package we used only supports the use of Gaussian PDF to estimate continuous variables, we log-transformed the preceding pauses. Finally, the task we are modeling is interactive, meaning that the state of the environment changes both because of actions taken by the student and actions taken by the interface. We introduce a third data stream that codes the actor producing the action.

We used the *HeterogeneousHMM* package [9] to estimate five HHMM models using the three data streams generated by 265 students who completed the science ICT. We considered models with between 8 and 12 states, using Bayesian Information Criterion to determine the best fitting model.

## 4. Results

We identified that a model with 11 latent states best fit the data. Table 1 shows the average response latency for these latent states and the probability these states are generated by students or the system. Our HHMM was able to distinguish:

a. States related to setting up and running experiments (i.e., States 6, 9, 10, 11)
b. Evaluating evidence using data tables states (i.e., States 2, 3),
c. A state for answering questions (i.e., State 1),
d. A state that appears to be a more meta-cognition related state (i.e., State 5).

We also distinguish between thoughtful and faster states (e.g., State 2 vs. State 3; State 9 vs. State 6).

## 5. Conclusions and Practical Implications

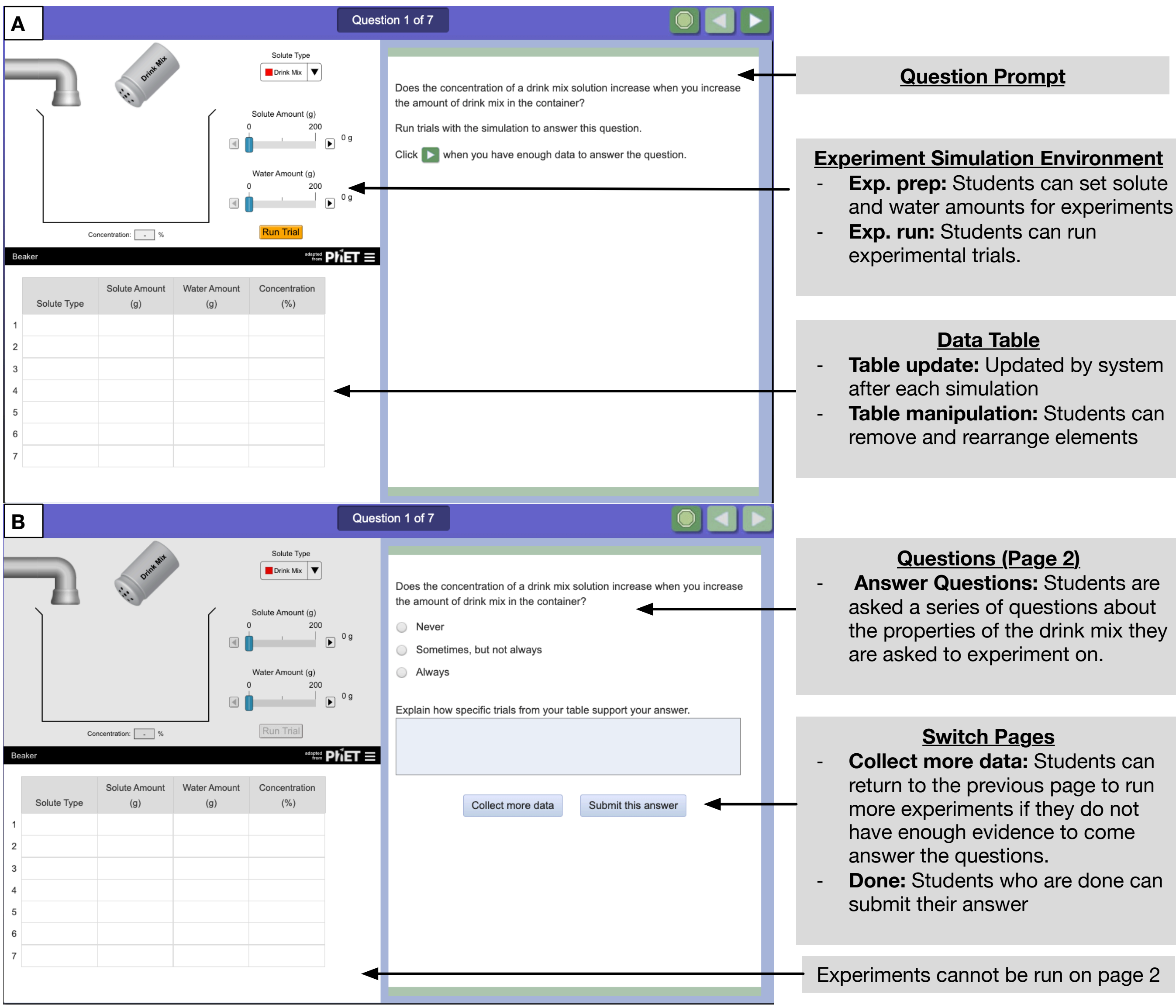
While response latency has a long history of use within cognitive science and assessment, the use of response latency within data driven models of task execution is limited [11]. This work presents, early evidence that HHMMs offer an effective descriptive approach for capturing latent problem-solving states while accounting for both the context and timing of actions. In future work we will explore how this method can be combined with clustering approaches to distinguish strategy use and student skill. Beyond the descriptive utility of these types of models (e.g.,[6]), HHMMs are a useful tool for building adaptive tasks, where increased probability of being in certain states can inform how content is delivered, hints are provided, or items are scored.

## 6. References


1. Anderson, J. R., & Fincham, J. M. (2014). Extending problem-solving procedures through reflection. *Cognitive psychology*, *74*, 1-34.
2. Paquette, L., de Carvahlo, A., Baker, R., & Ocumpaugh, J. (2014, July). Reengineering the feature distillation process: A case study in detection of gaming the system. In *Educational data mining 2014*.
3. Arslan, B., Lehman, B., & Lindner, M. (July, 2021). Introducing a theory-driven method to detect different levels of engagement in technology-enhanced items. *Presented at the 12th Conference of the International Test Commission, Virtual*.
4. Lehman, B., & Arslan, B. (July, 2021). Test-taker engagement levels when responding to drag-and-drop mathematics items. *Presented at the 12th Conference of the International Test Commission*, Virtual.
5. Lee, Y. H., & Jia, Y. (2014). Using response time to investigate students' test-taking behaviors in a NAEP computer-based study. *Large-scale Assessments in Education*, *2*(1), 1-24.
6. Bergner, Y., & von Davier, A. A. (2019). Process data in NAEP: Past, present, and future. *Journal of Educational and Behavioral Statistics*, *44*(6), 706-732.
7. Arieli-Attali, M., Ou, L., & Simmering, V. R. (2019). Understanding test takers' choices in a self-adapted test: a hidden Markov Modeling of process data. *Frontiers in psychology*, *10*, 83.
8. Xiao, Y., He, Q., Veldkamp, B., & Liu, H. (2021). Exploring latent states of problem-solving competence using hidden Markov model on process data. *Journal of Computer Assisted Learning*.
9. Moreno-Pino, F., Sukei, F., and Artes-Rodrigez, A. (2020). First stable release of HeterogeniousHMM. https://github.com/fmorenopino/HeterogeneousHMM
10. Schunn, C. D. & Anderson, J. R. (1998). Scientific discovery. In J. R. Anderson, & C. Lebiere (Eds.). *The Atomic Components of Thought*, 385-428.
11. Tenison, C., Arslan, B.: Characterizing pause behaviors in a science inquiry task. In:

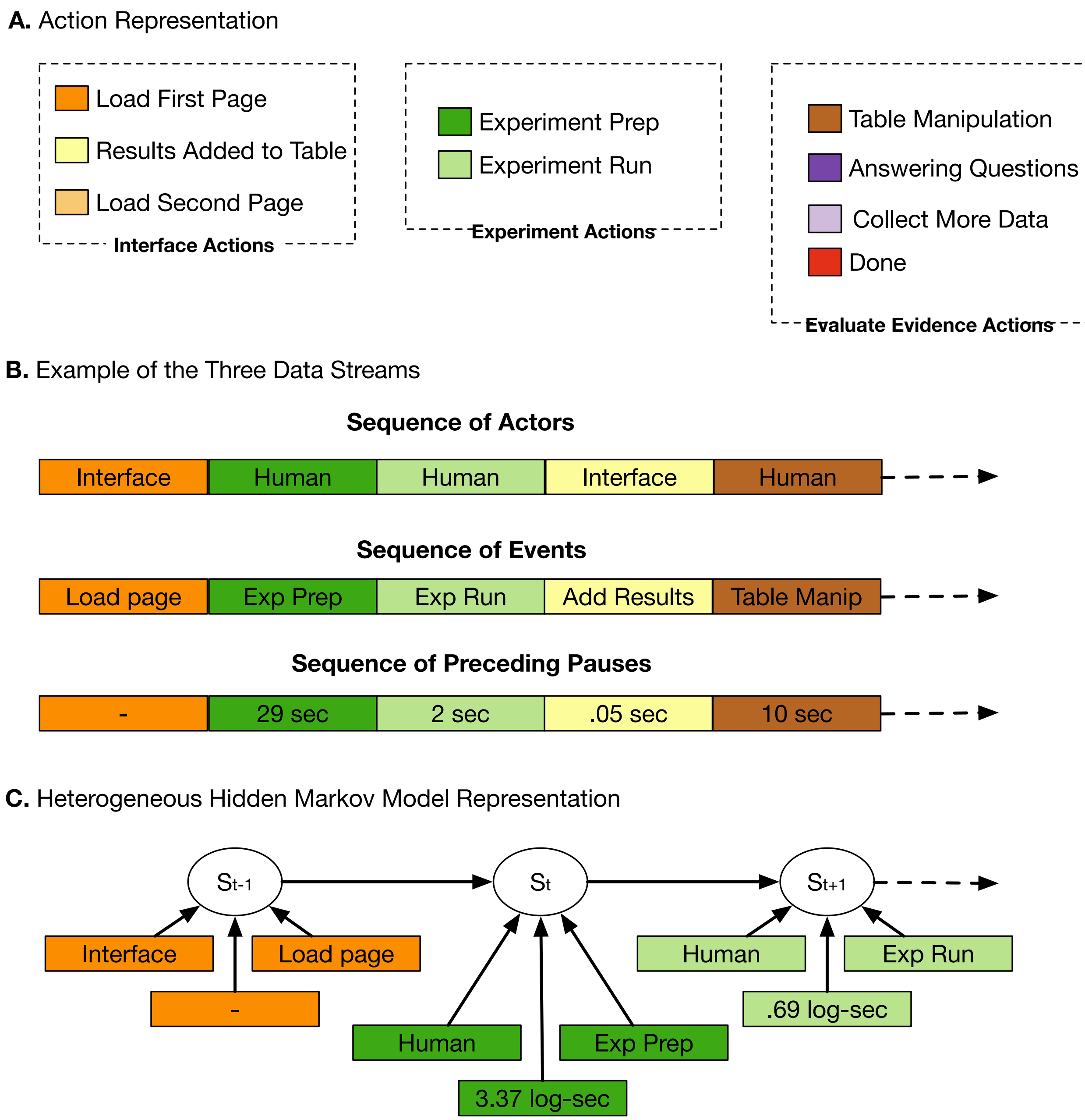
Stewart, T.C. (Ed.). Proceedings of the 18th International Conference on Cognitive Modeling, Applied Cognitive Science Lab, Penn State, University Park, PA, pp. 283–298 (2020)

**Figure 1**.




*Note.* A screenshot of question 1 in the concentration simulation: a) Science inquiry screen, b) Answer screen

*Figure 2*




*Note.* **A.** Nine Actions that occurred during the task grouped according to whether those actions were issued by the interface and the type of goals those actions helped students achieved. **B.** An example visualization of what the three data streams we modeled with our HHMM. **C.** An example of how the data-streams were used by the HHMM to estimate the probability of State *S* at time *t*.

**Table 1.**

*Descriptive information about the 11 states fit by our HHMM.*

| State | Gaussian of Log-transformed Preceding Pauses (log-sec) | | Probability of Actor | | Qualitative State Coding |
|---|---|---|---|---|---|
| | Mean | Scale | Student | Interface | |
| 1 | 1.57 | 1.37 | 1 | 0 | **Evaluate evidence:** Answer questions |
| 2 | 0.45 | 1.14 | 1 | 0 | **Evaluate evidence:** Thoughtful table manipulation |
| 3 | -5.41 | 2.50 | 1 | 0 | **Evaluate evidence:** Table manipulation |
| 4 | -0.98 | 2.31 | 0 | 1 | Task environment runs experimental simulation for trials with small amounts of solute and solvent and updates table when complete |
| 5 | 3.29 | 1.56 | 1 | 0 | **Meta-cognitive:** Students in this state reflect on the current goal and set high-level plan for next goal. These goals reflect deciding to run an experiment, manipulate data, answer questions, submit response |
| 6 | -0.59 | 0.34 | 1 | 0 | **Experiment:** Experiment set up (fast) |
| 7 | -9.21 | 0.00 | 0 | 1 | Simulation environment loads questions |
| 8 | 0.90 | 0.02 | 0 | 1 | Task environment runs experimental simulation for larger amounts of solute and solvent and updates table |
| 9 | 1.21 | 0.40 | 1 | 0 | **Experiment:** Thoughtful experiment set up |
| 10 | -1.61 | 0.05 | 1 | 0 | **Experiment:** Experiment set up (double click) |
| 11 | 0.56 | 0.83 | 1 | 0 | **Experiment:** Decide to run experimental simulation |

*Note.* We report, mean and scale of the Gaussian probability density function (PDF) estimated for emissions of the log-transformed preceding pauses, alongside the probability each state reflects a human action or an interface action. Action emission probabilities are captured in Figure 3.

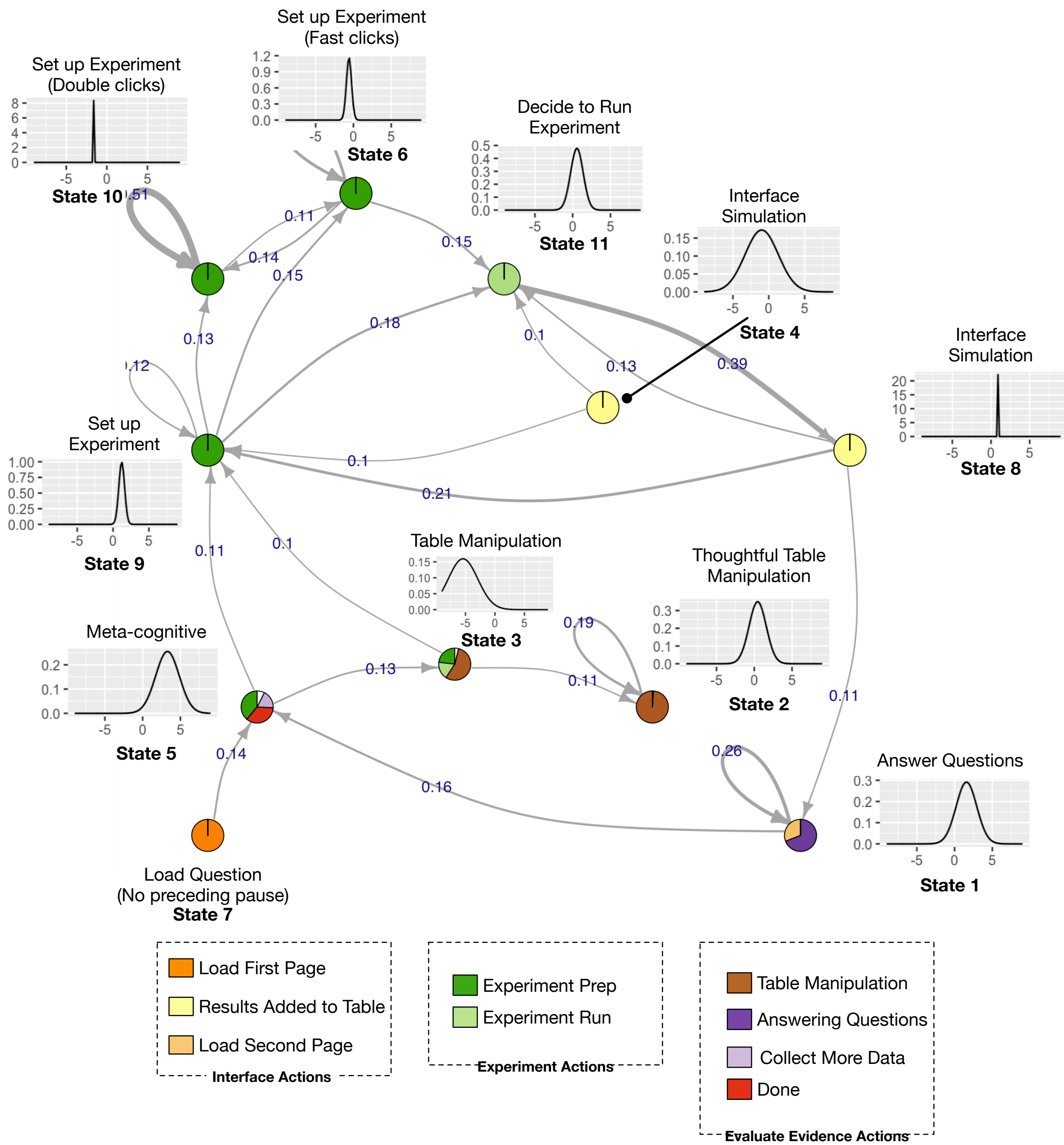



**Figure 3.** Node and arrow representation of HHMM fit to data streams capturing the action taken, the actor who took the action, and the pause preceding the action. Nodes represent hidden states with the color reflecting the probability of the hidden state emitting action events (color coded in the legend along the bottom). Arrows represent the transition probabilities, with labels and density reflecting specific probabilities. For readability we do not display transition probabilities less than .099. The Gaussian probability density functions graphed by each node represent the functions our model estimated to capture the pause latencies (log-sec) emitted by the 11 hidden states.