

# Simulating Multi-Subject Momentary Time Sampling

Luc Paquette

Teachers College, Columbia U.  
525 W. 120<sup>th</sup> Street  
New York, New York 10027  
+1 212 678 3854  
paquette@tc.columbia.edu

Jaclyn Ocumpaugh

Teachers College, Columbia U.  
525 W. 120<sup>th</sup> Street  
New York, New York 10027  
+1 212 678 3854  
jo2424@tc.columbia.edu

Ryan S. Baker

Teachers College, Columbia U.  
525 W. 120<sup>th</sup> Street  
New York, New York 10027  
+1 212 678 8329  
baker2@exchange.tc.columbia.edu

## ABSTRACT

This paper presents software for examining measurement error in Momentary Time Sampling—an interval time sampling method commonly used in research domains (e.g., classroom observations) where continuous recording is not feasible. The Parameters for Optimizing Scientific Sampling Using Momentary-time-sampling Simulator (POSSUMS) produces Monte Carlo simulations (based on user-specified values) and automatically generates statistics relevant to understanding the extent to which measurement error may be expected within multi-subject design parameters.

## Keywords

Momentary Time Sampling, Monte Carlo simulation, student behaviors, classroom observations.

## 1. INTRODUCTION

Educational research and other investigations of behavior often rely on sampling procedures when continuous observation is not viable [4]. As researchers in Educational Data Mining (EDM) have sought training labels for affect/engagement detectors to study the effects of student classroom behaviors on long-term outcomes, they have also relied on sampling procedures (e.g., [6]’s review). These include momentary time sampling (MTS), where researchers divide the observation session into intervals, recording whether a particular behavior occurs at the end of each. MTS, also known as *instantaneous time sampling* or *point sampling*, proves more accurate than similar techniques, including whole interval recording (WIR, where behavior is only recorded if it was present throughout the sampling interval) or partial interval recording (PIR, where behavior is recorded as present if it occurs at any time during the interval) [8]. Still, MTS is prone to substantial measurement error for some study designs [5].

Measurement error in MTS—the difference between *actual* and *observed* values for specific behaviors—is influenced by a large number of interacting factors [14], but research focuses on the duration of the behavior being observed and the length of the observation interval (e.g., [1]). Although the method does not introduce bias, the sometimes substantial variation in apparently transient measurement error has led to highly conservative suggestions, including [11], who suggest that MTS should only be used after continuous observations first determine typical values

for factors known to influence MTS measurement error.

A different approach to dealing with the uncertainty in MTS is to model measurement error through simulation (see extensive review in [14]), sometimes to study particular conditions and other times to make more general recommendations (e.g., [10], [13], [12], [3]). However, existing simulators [7] have focused on single-subject designs, which is inadequate for modeling measurement error in observation systems where an observer is coding multiple students in the same session (e.g., BROMP [6], a common method for EDM research; but also classroom observation schemes used by many public schools in the U.S.). In this poster, we present a freely available simulator that addresses this gap: the Parameters for Optimizing Scientific Sampling Using Momentary-time-sampling Simulator (POSSUMS).

## 2. Prior Research

Prior research has shown that measurement error in MTS may be induced by a number of interacting factors, including: (a) the *sampling interval* (how often observations are recorded) (b) the *observation session’s length*, (c) *bout-length* (the duration of each event/behavior being observed) and (d) *prevalence*, the percentage of an observation session that a behavior occupies (as [6] and [14] review). Previous research with simulations has led to practical recommendations, such as specific limits on sampling intervals (e.g., less than every 60 seconds [2] or 120 seconds [9]), or more general suggestions (e.g., sampling intervals must be shorter than mean bout-length [1], [14]), but these recommendations are based on simulations involving single-subject design. That is, these are recommendations for estimating the amount of time that a single research subject (e.g., a student) spends engaging in a particular behavior (e.g., on-task conversation) over a given observation session (e.g., an hour long class). They have not been demonstrated to be appropriate for estimating prevalence in multi-subject research designs (e.g., the amount of time that students in a particular classroom spend engaged on on-task conversation over the course of a class session). What’s more, simulators that are currently publically available for single-subject design (e.g., [7]) require programming skills, limiting their use to researchers familiar with that programming language.

## 3. POSSUMS 1.0

In this paper, we present POSSUMS 1.0: a java-based tool that allows researchers using MTS in multi-subject design to run Monte Carlo simulations to study potential measurement error. POSSUMS, which is freely available on the 1st author’s webpage (<http://www.columbia.edu/~lp2575/tools.html>), allows users to set parameters which it models, automatically generating metrics needed to understand potential sources of error. In the sections that follow, we present the user interface and the output.

### 3.1 User Interface

POSSUMS presents users with an interface that allows them to set several relevant parameters. As shown in Figure 1, users first add

target behaviors to be observed, specifying projected bout-length and prevalence for each. They then specify how many subjects (students) will be coded and the length of the observation window. (These values are used to run a Monte Carlo simulation that represents the *actual*, continuous values that might be found *en vivo*.) The user also specifies multiple sampling intervals, in seconds, which are used to simulate MTS *estimates*—the values that would be obtained based on sampling at those intervals across multiple subjects. Finally, the user indicates how many simulations should be run.

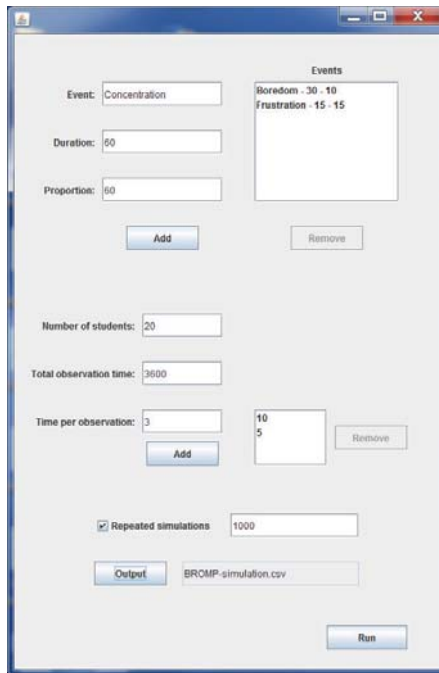


Figure 1. POSSUMS 1.0 User Interface

### 3.2 Output

POSSUMS 1.0 outputs to .csv files, which import easily into Excel or other widely-used analyses tools. The exact format depends on how many simulations are run. When only a single simulation is executed, the output file summarizes how many times each target behavior was observed, providing percentages for each behavior's contribution to the total observations at the classroom and the student level. These files also contain a detailed list of the behaviors associated to each student at each second in the simulated observation period. When multiple simulations are executed, the output file is different, providing summary measures that average across all simulations. Those values include the average and standard deviations for the number of times the behavior was observed across simulations, and the average and standard deviation for the percentage of observations for each behavior across simulations.

### 4. Discussion/Conclusions

Educational research, like other domains that sometimes require observational research, has long relied on sampling methods to estimate actual values. As EDM research begins to make use of observational methods to estimate the prevalence of relevant behaviors or events in classroom settings (cf. [6]), it is important that researchers understand possible sources of measurement error. Because this error in MTS appears to be influenced by a large number of factors working in concert, to date efforts to quantify it have focused on single-subject design (e.g., [7], [14]).

Unfortunately, these studies are insufficient for understanding measurement error in many observational studies of classroom conditions, which involve multi-subject designs. POSSUMS 1.0 represents a step forward in this effort, simulating pertinent field conditions and automatically generating metrics needed to understand potential sources of error.

### 5. ACKNOWLEDGMENTS

Thanks to James Pustejovsky, Elizabeth Tipton, Didith Rodrigo, and Sweet San Pedro, for help understanding issues presented here. Special thanks to Stefan Slater, who motivated the name.

### 6. REFERENCES

- [1] Ary, D., & Suen, H. K. 1983. The use of momentary time sampling to assess both frequency and duration of behavior. *Journal of Behavioral Assessment*, **5**, 143–150.
- [2] Brittle, A. R., & Repp, A. C. 1984. An investigation of the accuracy of momentary time sampling procedures with time series data. *British Journal of Psychology*, **75**, 481–488.
- [3] Fiske, K., Delmolino, L. 2012. Use of discontinuous methods of data collection in behavioral intervention: Guidelines for practitioners. *Behavior Analysis in Practice*, **5**, 77–81.
- [4] Mudford, O. C., Taylor, S. A., & Martin, N. T. 2009. Continuous recording and interobserver agreement algorithms reported in the *J. of Applied Behavior Analysis* (1995–2005). *J. Applied Behavior Analysis*, **42**, 165–169
- [5] Murphy, G., & Goodall, E. (1980). Measurement error in direct observations: A comparison of common recording methods. *Behaviour Research & Therapy*, **18**, 147–150.
- [6] Ocumpaugh, J., Baker, R., Rodrigo, M. 2015. *Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 Tech. & Training Manual*. NY, NY: Teachers College, Columbia U. Manila, PH: Ateneo Laboratory for the Learning Sciences.
- [7] Pustejovsky, J. E., & Runyon, C. 2014. Alternating Renewal Process Models for Behavioral Observation: Sim.Methods, Software, & Validity Illustrations. *Behav. Disorders*, **39**(4).
- [8] Rapp, J. T., Colby-Dirksen, A. M., Michalski, D. N., Carroll, R., & Lindenberg, A. 2008. Detecting changes in simulated events using partial-interval recording & momentary time sampling. *Behavioral Interventions*, **23**, 237–269.
- [9] Rhine, R. & Ender, P. 1983. Comparability of methods used in sampling primate behavior. *Am. J. Primatology*, **5**, 1–15.
- [10] Rojahn, J., & Kanoy, R. 1985. Toward an empirically based parameter selection for time-sampling observation systems. *J. Psychopathology & Behavioral Assessment*, **7**, 99–120.
- [11] Sanson-Fisher, R., Poole, A., & Dunn, J. 1980. An empirical method for determining an appropriate interval length for recording behavior. *J. App. Behavior Analysis*, **13**, 493–500.
- [12] Suen, H., & Ary, D. 1986. A post hoc correction procedure for systematic errors in time-sampling duration estimates. *J. Psychopathology & Behavioral Assessment*, **8**, 31–38.
- [13] Wilson, R., Jansen, B., & Krausman, P. 2008. Planning & assessment of activity budget studies employing instantaneous sampling. *Ethology*, **114**, 999–1005.
- [14] Wirth, O., Slaven, J., & Taylor, M. 2014. Interval sampling methods and measurement error: A computer simulation. *Journal of Applied Behavior Analysis*, **47**(1), 83-10.