Minimizing and Reporting Momentary Time-Sampling Measurement Error in Single-Case Research

Kathleen B. Cook^{1,2} · Sara M. Snyder^{1,3}

Published online: 17 January 2019
© Association for Behavior Analysis International 2019

Abstract

Research indicates that momentary time sampling (MTS) is often the best interval-measurement system when observing duration of behavior. Several recent studies recommended considering mean duration of target behavior, as well as durations of measurement intervals and observation sessions, to minimize measurement error in MTS. This report describes the steps we used to minimize measurement error in a single-case design research study. Further, we detail our methods for monitoring and reporting MTS measurement error across conditions by intermittently collecting and analyzing duration per occurrence measurements.

Keywords Interval recording · Momentary time sampling · Measurement error · Single-case research design

Interval recording systems present a desirable alternative to continuous measurement systems. They are frequently used in educational settings because they require less time and resources (Lane and Ledford 2014). However, because not every behavior event or duration is recorded, interval systems provide only an estimate of the observed behavior. Having familiarity with the deficiencies of interval measurement systems will aid researchers using single-case designs in implementing the most appropriate interval measurement system with minimal systematic error (Ledford et al. 2015).

Of the three commonly used interval systems, momentary time sampling (MTS) is commonly thought to result in less measurement error than whole-interval recording (WIR) and partial-interval recording (PIR; Fiske and Delmolino 2011). Like PIR and WIR, MTS is an interval recording system used to estimate the count or duration of behavior. The observation session is first divided into intervals, and then at the end of each interval, the observer codes the occurrence or

nonoccurrence of target behavior (Ayres and Ledford 2014). Research using computer simulations has shown MTS to be unbiased when estimating absolute error, whereas other time-sampling methods consistently over- or underestimate absolute error (WIR and PIR, respectively; Wirth et al. 2014).

Although additional research is needed, some guidelines are emerging to support the selection and design of a specific interval measurement system in consideration of variables such as interval and session duration (Fiske and Delmolino 2011). Research indicates that, for the three interval systems, the shorter the interval, the lower the measurement error (Fiske and Delmolino 2011; Wirth et al. 2014). Additionally, a recent study of computer-simulated interval measurements (Wirth et al. 2014) concluded that less error occurred when interval lengths approximated or were shorter than the event being recorded. Sharp et al. (2015) reported more representative estimates with shorter MTS intervals, longer observation sessions, and longer behavior durations. A summary of current literature on interval measurement systems used in early childhood special education and early intervention research (Lane and Ledford 2014) found that MTS was generally accurate when intervals of less than 1 min were used—and with intervals less than 30 s producing the most accurate estimates.

Further, Sharp et al. (2015) used computer simulation to evaluate data obtained from the applied setting, analyzing data of varying bouts and total duration to develop decision rules for selecting MTS interval length. Sharp et al. recommended collecting preintervention data to determine the initial MTS interval, based on an approximation of the mean bout of

✉ Kathleen B. Cook
kathleen.cook@augie.edu

¹ Department of Communication Sciences and Special Education, University of Georgia, Athens, GA, USA

² Education Department, Augustana University, 2001 South Summit Avenue, Sioux Falls, SD 57197, USA

³ Educational Foundations and Exceptionalities Department, James Madison University, Harrisonburg, VA, USA

behavior. Additionally, Ledford et al. (2015) recommended that, when reporting MTS data, researchers also provide an estimate of measurement error by presenting average duration per occurrence (DPO) in some sessions in every condition and for every participant.

In addition to interval lengths, the duration of the observation session should also be considered when attempting to minimize measurement errors with MTS. Research indicates that longer observation sessions using interval measurements more closely approximated the data from continuous measures, especially with lower occurring behaviors (Fiske and Delmolino 2011). Similar results were reported with computer simulations of interval measurements, in which longer observation sessions were correlated with less error (Wirth et al. 2014).

Another recent simulation study (Ledford et al. 2015) evaluated interval measurement systems and concluded that continuous rather than discontinuous recordings were preferred for short observation sessions or when measuring frequency of behavior. However, similar to previous research, the authors credited MTS with providing the best approximation of behavior duration. Additionally, the authors recommended that researchers using MTS conduct DPO measurement sessions during baseline and intervention conditions of a study to provide an estimate of the error derived from the use of an interval measurement system. Changes should be made to the interval length across conditions, if warranted, as DPO may change in response to intervention (Ledford et al. 2015).

The purpose of this report is to describe the steps taken during a recent study to address measurement error when using an interval-based system for estimating behavior duration. We used MTS to measure the effects of a self-monitoring strategy used by high school students with disabilities on classroom engagement (i.e., on-task) behavior. We describe steps taken a priori to select appropriate interval and session durations in order to minimize measurement error, and we recount the research plan to monitor measurement error by intermittently collecting DPO data. Using a manual conversion of DPO to MTS, we also report the measurement error that would have resulted from using MTS on sessions in which DPO was used.

Method

Participants and Setting

Teachers were asked to nominate students who demonstrated low levels of on-task behavior in math class. Four 9th- and 10th-grade students with high-incidence disabilities participated. The study took place in two mathematics classrooms with two participants per classroom.

Observers and Data Collectors

The primary observer during all conditions of the study, the first author, was completing doctoral studies in special education and working toward certification in behavior analysis (Board Certified Behavior Analyst–Doctoral, BCBA-D). Three doctoral students in special education were trained as reliability observers by the first author; two of the reliability observers were BCBAs and the third was studying behavior disorders. All observers had prior experience as special education teachers.

Measuring On-Task Behavior

Definitions for on-task behavior were developed in consultation with the teachers. For most sessions, the data-collection system used by the researchers was MTS; however, DPO was intermittently collected instead of MTS. To collect MTS data, observers checked behavior at the end of intervals, recording only the occurrence (+) or nonoccurrence (–) of behavior at that moment. An interval timer on a data-collection smartphone app signaled 10-s intervals, allowing us to observe the two participants at 20-s intervals. The app, ABC Data Pro (Romanczyk and Gillis n.d.), was designed for behavior analysts and special educators and featured several forms of continuous and noncontinuous data collection. According to ABC Data Pro's webpage (<http://cbtaonline.com/drupal/abcdatapro>) at the time of the study, the app had not been updated for newer versions of Apple devices (i.e., iOS 7 and iOS 8). Observations were rotated across participants; thus, with two participants in a classroom, each participant was observed at 20-s intervals for 30 min. After each session, researchers calculated the percentage of total on-task intervals for each participant by dividing the number of on-task intervals by the total number of intervals (i.e., the number of 20-s intervals observed for that participant) and multiplying by 100, providing an estimate of the overall occurrence of on-task behavior.

Instead of MTS—in 24% of baseline and in 10% of intervention sessions—DPO data were captured and recorded via the same data-collection app on the smartphone. In order for one observer to collect DPO data on two participants within a session, we resorted to a noncontinuous but sequential observation system that allowed us to determine mean durations of the observed behaviors (Thomson et al. 1974). One participant was observed over a 5-min segment, and then the other participant in the same classroom was observed for the next 5-min segment; the rotation of observations continued for the session. Each participant was observed for a total of 20 min during the session. Next, we calculated the mean bout of observed behaviors, which we used to determine whether our MTS intervals were appropriate (i.e., intervals approximated or were shorter than the event being recorded; Wirth et al. 2014).

Our last step was to manually convert each participant's DPO from the four 5-min observations (staggered with the observations of the other participant in the classroom) to represent MTS data for analysis of measurement differences. DPO data from ABC Data Pro provided information on the starting and ending times of each event recorded during the observation window. To manually convert to MTS, the starting and ending times of each DPO event were aligned on a spreadsheet alongside a column of 20-s intervals. A DPO event that intersected with the end of any 20-s interval was marked with a 1, and remaining 20-s intervals were marked with a 0 (indicating nonoccurrences of the target behavior). The total number of 20-s intervals in which a DPO event was recorded was then divided by the total number of 20-s intervals in the observation session to determine the equivalent MTS percentage. Figure 1 and Table 1 provide comparisons of DPO data with MTS data (that were converted from DPO).

Experimental Design

We used a counterbalanced, multiple-treatment single-case research design (Harris et al. 2005). The multiple-treatment design allowed comparison of two self-monitoring intervention conditions, with and without self-graphing.

Procedures to Minimize and Monitor Measurement Error

Initial Selection of Interval Length We used sequential 5-min observations (Thomson et al. 1974) to determine mean bouts of behavior in the initial baseline session for all participants. Then we consulted Table A1 from Wirth et al. (2014), which provided examples of the absolute error that would be associated with various interval durations in a time-sampling system, based on the event duration and

Fig. 1 Graphs showing percentages of on-task behavior with momentary time-sampling (MTS) and duration per occurrence (DPO) measures. Open circles indicate MTS data, and bars indicate DPO data (in sessions in which DPO data were collected, data were converted to MTS; results from both recording methods are shown); A = baseline; B = self-monitoring; C = self-monitoring plus self-graphing

Table 1 DPO With Conversion to MTS and Error Values

	Baseline 1	Baseline 2	Intervention 1	Intervention 2
DPO Mean %	44.78	39.96	63.74	65.38
MTS Mean %	46.37	41.00	65.00	64.33
Minimum/Maximum Error	−0.92/7.33	−1.50/3.27	−3.68/6.41	−4.02/1.67

Positive error values represent overestimations, and negative error values represent underestimations

Intervention 1, self-monitoring; *Intervention 2*, self-monitoring plus self-graphing; *DPO*, duration per occurrence; *MTS*, momentary time sampling at 20-s intervals

the percentage of actual cumulative event duration over a 1-h observation period. Wirth et al.'s simulations indicated that, with event durations between 1 s and 256 s for behaviors occurring between 20% and 80% of a 1-h observation session and sampling intervals between 15 s and 30 s, estimated measurement error would be lowest (i.e., projected median error values between −0.3 and 0.3). Using the 5-min sequential observations within total session observations of 20 min per participant, the mean bouts of behavior observed in the first baseline sessions across participants ranged from 21.96 s to 32.79 s, occurring at a mean percentage of 53% of the observation window (range 35%–66%). We had already considered using a 20-s MTS interval, which would require the observer of two participants to observe and record data at 10-s intervals. Thus, with the initial session of baseline data indicating the lowest mean duration of 21.96 s of on-task behavior across participants, the appropriateness of our selected 20-s observation interval was confirmed.

Selection of Session Duration Recent research indicated that longer sessions of interval sampling would yield more representative duration data (Sharp et al. 2015; Wirth et al. 2014). Therefore, with daily class times ranging between 40 and 55 min, we planned to conduct MTS observations during 30-min sessions. Observations and data collection commenced approximately 5 min after the class bell rang.

Monitoring Event Duration and Measurement Error To provide an ongoing estimate of MTS measurement error, we planned to collect DPO data intermittently, at least twice during baseline and once during intervention for each participant. Our method of determining mean DPO was through sequential 5-min observations across two participants within a session (Thomson et al. 1974). We hypothesized that on-task bouts and overall percentages of on-task behavior would increase in intervention conditions, further minimizing the measurement error (Sharp et al. 2015). Therefore, we planned to monitor data and consider changing interval length only in cases of nonresponse to intervention in which the mean DPO of on-task behaviors became smaller than the 20-s MTS interval.

Results and Discussion

Interobserver Agreement

A second observer collected interobserver agreement (IOA) data on a minimum of 20% of MTS sessions and 36% of DPO sessions. Mean IOA for on-task behavior in MTS sessions was calculated using three methods (Ayres and Ledford 2014): overall mean agreement was 93% when using the interval-by-interval method (range 83%–97%), 89% when using occurrence (range 65%–95%), and 80% when using nonoccurrence (range 33%–94%). The mean total duration IOA for on-task behavior in DPO sessions was 89% (range 82%–96%).

Estimating MTS Measurement Error

To provide an ongoing estimate of MTS measurement error, DPO data were collected in 24% of baseline sessions and 10% of intervention sessions. Figure 1 depicts on-task percentages for each participant. In DPO sessions, both DPO and converted MTS estimates of the percentages of on-task behaviors are shown on the line graphs. The reader should be cautioned that the DPO data were collected in a noncontinuous sequential method (with 5-min segments of sequential observation across two participants per session) and, therefore, represent only an estimate of the total duration of on-task behavior during the session. Additionally, Table 1 shows the mean participant DPO and converted MTS percentages along with minimum and maximum error values.

The overall mean DPO of all participants from sessions in the initial baseline condition was 44.76%, but when converted to MTS data at 20-s intervals, the mean was slightly higher, at 46.85%. The overall range of error values (−4.02 to 7.30) was smaller than the range estimates from Wirth et al. (2014; −14.8 to 14.8); however, an exact comparison was not available from Table A1 in Wirth et al., which only provided error estimates for 15-s and 30-s intervals across 60-min observation periods, whereas the intervals in this study were 20 s with an observation session of 30 min. As expected, on-task bouts and overall percentages of on-task behavior increased in

intervention conditions (range 23.35–127.30 s), so we maintained MTS intervals of 20 s throughout the study.

Summary of Findings

This study employed a multiple-treatment single-case research design to evaluate the effects of self-monitoring treatments on the on-task behavior of high school students with disabilities in general education classes. We addressed the errors inherent in a discontinuous measurement system, MTS, in two ways: (a) using initial-session baseline measures of event duration and overall percentage of the event within the observation session, along with Wirth et al.'s (2014) error tables, to select a time-sampling interval that would result in minimal measurement error, and (b) collecting DPO (using 5-min sequential observations across two participants per session) instead of MTS data during several sessions in baseline and intervention conditions to check the appropriateness of the MTS observation interval selected, as recommended by Ledford et al. (2015). Our analysis indicated time-sampling errors ranged from underestimates of 4.02% to overestimates of 7.30%. Next, we graphed DPO with MTS data to illustrate potential measurement errors in MTS. Visual analysis of graphed data showed MTS data were similar to DPO and appeared to confirm our application of interval-selection guidelines.

Implications for Practitioners and Researchers

Although total-duration measurements will be most accurate, an interval measurement system may often be more practical and efficient. For example, when recording DPO, we were able to observe only one participant at a time. However, with MTS intervals of 20 s, we collected data on two participants, rotating observations at 10-s intervals, allowing time between intervals to record and prepare for the next observation. In practice, the behavior analyst or classroom teacher collecting data on a behavior with a similar DPO would likely find it more practical to observe only one individual at 20-s intervals.

Additionally, future research on identifying factors and solutions related to measurement error in interval time-sampling systems is warranted. Specifically, as recommended by Ledford et al. (2015), future studies should continue to establish quantifiable relationships between DPO and interval sizes so that the practitioner can reliably select the most practical (i.e., longest) interval that will result in accurate estimates of behavior. Research should include investigations on the relations between the session duration, the DPO of the target behavior, and optimal intervals for observation (Sharp et al. 2015).

However, until more effective data-based decision tools have been developed (Sharp et al. 2015; Wirth et al. 2014), we recommend that researchers and practitioners consider current time-sampling research when selecting interval method,

interval duration, and observation session. Further, we recommend following the recommendations of Ledford et al. (2015) to use DPO to estimate and report errors in MTS data. Additionally, although continuous-duration measurements provide the most accurate data on the percentage of behavior during a session, a noncontinuous sequential observation as described by Thomson et al. (1974) allows for the calculation of mean DPO when the observer is measuring the behaviors of more than one student. When using MTS in research and applied practice, intermittently collecting DPO data, using the mean bout of behavior to determine the appropriateness of observation intervals, and reporting potential measurement error of MTS will strengthen researcher and practitioner confidence in the data.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Ayres, K., & Ledford, J. R. (2014). Dependent measures and measurement systems. In D. L. Gast & J. R. Ledford (Eds.), *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed., pp. 124–153). New York, NY: Routledge.
- Fiske, K., & Delmolino, L. (2011). Use of discontinuous methods of data collection in behavioral intervention: Guidelines for practitioners. *Behavior Analysis in Practice*, 5(2), 77–81. <https://doi.org/10.1007/BF03391826>.
- Harris, K. R., Friedlander, B. D., Saddler, B., Frizzelle, R., & Graham, S. (2005). Self-monitoring of attention versus self-monitoring of academic performance: Effects among students with ADHD in the general education classroom. *Journal of Special Education*, 39(3), 145–156. <https://doi.org/10.1177/00224669050390030201>.
- Lane, J. D., & Ledford, J. R. (2014). Using interval-based systems to measure behavior in early childhood special education and early intervention. *Topics in Early Childhood Special Education*, 34(2), 83–93. <https://doi.org/10.1177/0271121414524063>.
- Ledford, J. R., Ayres, K. M., Lane, J. D., & Lam, M. F. (2015). Identifying issues and concerns with the use of interval-based systems in single case research using a pilot simulation study. *Journal of Special Education*, 49(2), 1–14. <https://doi.org/10.1177/0022466915568975>.
- Romanczyk, R. G., & Gillis, J. M. (n.d.). ABC data pro [smartphone application]. CBATonline—Tools for behavior analysis. Retrieved from <http://cbatonline.com>

- Sharp, R. A., Mudford, O. C., & Elliffe, D. (2015). A data-based method for selecting parameters of momentary time sampling to provide representative data. *European Journal of Behavior Analysis*, 16(2), 279–294. <https://doi.org/10.1080/15021149.2015.1083294>.
- Thomson, C., Holmberg, M., & Baer, D. M. (1974). A brief report on a comparison of time-sampling procedures. *Journal of Applied Behavior Analysis*, 7, 623–626. <https://doi.org/10.1901/jaba.1974.7-623>.
- Wirth, O., Slaven, J., & Taylor, M. A. (2014). Interval sampling methods and measurement error: A computer simulation. *Journal of Applied Behavior Analysis*, 47, 83–100. <https://doi.org/10.1002/jaba.93>.