2011

An analysis of the methodological and human error within momentary time sampling data collection

Joslyn Cynkus Mintz
Louisiana State University and Agricultural and Mechanical College

AN ANALYSIS OF THE METHODOLOGICAL AND HUMAN ERROR WITHIN MOMENTARY TIME SAMPLING DATA COLLECTION

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in
The Department of Psychology

by
Joslyn Cynkus Mintz
B.S., University of Georgia, 2001
M.S., Georgia State University, 2008
August 2011

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Teachers often serve as data collectors for the problem behavior of referred students in their classrooms; yet, the accuracy of teacher data collection has rarely been directly assessed. Momentary time sampling (MTS) may be a potentially useful option for teacher data collection because it does not require continuous monitoring, but rather requires the teacher to score the occurrence or non-occurrence of targeted behaviors at given instances. Research has shown that the smaller the interval between observations, the less methodological error will be introduced into MTS. However, the use of short-interval windows requires additional effort on the part of the teacher, and data collection becomes potentially more susceptible to competition with the teacher's other responsibilities. It is not clear based upon previous research to what extent human error influences the accuracy of MTS data. Therefore, the purpose of the current study was to evaluate the amount of methodological, human, and total error introduced during MTS data collection within two highly controlled experimental contexts, in which the duration of occurrence was determined. In highly controlled settings, results demonstrated that the amount of methodological and total error tended to increase as the MTS interval became longer and that human error was observed to be low across all MTS intervals.

INTRODUCTION

Direct observation and measurement of behavior, as opposed to relying on verbal reports, is a central tenet of applied behavior analysis as a science and as a practice (Baer, Wolf, and Risley, 1968). From a scientific perspective, high-quality measurement is essential to demonstrating orderly functional relations between a behavior and an environmental accompaniment (Johnston & Pennypacker, 1980). From a practical perspective, the efficacy of an intervention upon a socially important behavior will be determined in part by the employed measurement systems' capacity to detect changes in the levels of the targeted behavior. To the extent that a high-quality measurement system has been used, one can be reasonably confident in the accuracy of the conclusions drawn from one's data.

There are at least three factors that impact the quality of a measurement system from a scientific perspective. First and foremost, a measurement system needs to be accurate. That is, the obtained value of a behavior needs to correspond to the actual value of that behavior (Kazdin, 1982). Second, a measurement system needs to be reliable. That is, a measurement system needs to be capable of producing similar outcomes when applied repeatedly to the same phenomenon. Third, a measurement system must be capable of detecting changes in the level of the targeted behavior, often referred to as sensitivity. The more sensitive a measurement system is, the more accurate it is at detecting small changes in behavior (Johnston & Pennypacker, 1980).

When continuous data collection is used (frequency or duration recording), questions of the accuracy of data are minimized. That is, when each instance of a behavior is captured, one can be reasonably assured that the number recorded approximates the actual number of instances of the target behavior. Continuous data collection is often untenable and has led to the development of discontinuous measurement systems (Thomson, Holmberg, & Baer, 1974). As the name implies, discontinuous measurement involves collecting only a sample of the target behavior and assuming a representativeness of this sample. This measurement technique commonly involves dividing an observation period into equal intervals (e.g., a 15-min session could be divided into 90, 10-s intervals), and scoring the presence or absence of the target behavior during each interval. Discontinuous measurement, or time-sampling, confers a number of advantages over continuous data collection including the ease of use, the minimal amount of equipment needed, and the straightforwardness of observer training (Mann, Have, Plunkett, and Meisels, 1991). A number of different discontinuous measurement systems have been described in the literature, each with their own scoring rules.

When collecting data using whole-interval recording (WIR), a behavioral occurrence is scored only if the target behavior occurred for the *entire* duration of the interval. For example, Kazdin, Silverman, and Sittler (1975) used a 10-s WIR procedure to assess the vicarious effects of nonverbal attention on the attentiveness of a target student and adjacent peer within a special-education classroom. For each interval, the students were scored as being attentive if they were sitting in their seat, not talking to another student without permission, and working on the assigned task or paying attention to the teacher for the entire interval. Results demonstrated that nonverbal attention paired with a verbal prompt was effective in changing the behavior of both the target student and adjacent peer. In their investigation, WIR captured increases in attentiveness of these students following the introduction of positive reinforcement for attending.

When collecting data using partial-interval recording (PIR), a behavioral occurrence is scored if the target behavior occurred *at any point* within the interval. For example, Hall, Lund, and Jackson (1968) used a 10-s PIR procedure to assess the effect of contingent teacher attention on reducing off-task or disruptive behavior in elementary school children. For each participant, target behaviors were operationally defined and an interval was scored if there were any instances of off-task or disruptive behavior within each 10-s interval. Data were also collected on teacher behavior (verbalizations and proximity) using 10-s PIR by creating separate rows on the data sheet. Results demonstrated that when teacher attention was made contingent on appropriate study behavior, on-task behavior increased. The use of a PIR measurement system captured increases in on-task behavior for these students following the introduction of positive reinforcement for being on-task.

When collecting data using momentary-time sampling (MTS), a data collector only observes behavior at the very conclusion of an interval; a behavioral occurrence is scored if the target behavior occurs at the *exact moment* of observation. For example, Broden, Bruce, Mitchell, Carter, and Hall (1970) used a 10-s MTS procedure to record the on-task behavior of two students. During each 10-s interval, the on-task behavior of one student was recorded at the end of the first 5 s and the on-task behavior of the second student was recorded at the end of the next 5 s. Results demonstrated an increase of on-task behavior when it was followed by teacher attention. The results demonstrated that MTS was sensitive to changes in behavior.

As these discontinuous data collection systems collect only a sample of behavior, there are concerns regarding the extent to which those samples reflect the actual occurrence of the reported behaviors (i.e., the accuracy of discontinuous data collection systems; in particular from Johnson & Pennypacker, 1980; 1993). Powell, Martindale, and Kulp (1975) conducted a seminal study evaluating the accuracy of PIR, WIR, and MTS data relative to continuous data collection. In this study, the experimenters videotaped a secretary and collected data of her in-seat behavior during 20-min sessions; in-seat behavior ranged in occurrence from 10 to 90 percent of the session. The experimenters then scored each session using one of the discontinuous measurement systems and continuous (i.e., duration) measurement. Eight MTS intervals (i.e., 10, 20, 40, 80, 120, 240, 400, 600 s) and five WIR and PIR intervals (i.e., 10, 20, 40, 80, 120 s) were evaluated. Across each discontinuous measurement system, shorter intervals resulted in less error (i.e., were more similar to the continuous data collection system). The authors also reported that WIR and PIR were associated with characteristic forms of error. That is, WIR consistently underestimates the actual duration of behavior and PIR consistently overestimates the actual duration (see also Ary, 1984; Gardenier, MacDonald, & Green, 2004; Harrop & Daniels, 1986; Murphy & Goodall, 1979; Powell, Martindale, Kulp, Martindale, & Bauman, 1977; Rapp et al., 2008; and Saudargas & Zanolli, 1990). Interestingly, the data collection system in which the least amount of behavior is actually observed (i.e., MTS) was free of a predictable over or underestimation and provided the most accurate representation of in-seat duration up to intervals of 120 s (error increased as observations became more dispersed). These conclusions were somewhat limited in that the frequency and duration of the targeted behaviors varied across sessions and each session was scored using only one discontinuous measurement system.

Powell et al. (1977) addressed those limitations by systematically replicating Powell et al. (1975), but this time controlling the duration of the target behaviors. In-seat behavior was again measured, but in the current investigation the duration of in-seat behavior was manipulated to represent 20%, 50%, or 80% of the total 30-min session duration (6 min, 15 min, or 24 min, respectively). Videotapes of each session were scored using WIR, PIR, and MTS with 5-s intervals. Longer interval durations for each measurement system were then extracted from the data to represent longer intervals. Results were similar to Powell et al. (1975) in that MTS most closely corresponded to the actual durations of in-seat behavior; lacking the characteristic error of PIR and WIR. Further, minimal error was introduced to the accuracy by the MTS procedure up to intervals of 60 s; however error systematically increased as interval durations exceeded 60 s.

Research has also addressed the ability of MTS and PIR to accurately collect frequency data. Repp, Roberts, Slack, Repp, and Berkler (1976) compared two MTS procedures to two PIR methods and found MTS to be inaccurate in estimating rate of behavior. They used an event recorder to create data to represent behavior occurring at multiple rates of responding (i.e., high, moderate, and low) that occurred constantly across the session or in a burst pattern that resulted in a high rate of responding followed by a low rate of responding three times within a session. Results suggested that the MTS procedures produced relatively inaccurate estimates of behavior across the different rates of responding and interval recoding produced more accurate results for low and moderate response rates. However, it should be noted that in their MTS procedures the observations were spaced approximately 10 min apart, but the observations were made approximately every 10 s for the PIR methods; a more balanced comparison would use equal intervals. More recently Meany-Daboul, Roscoe, Bourret, and Ahearn (2007) compared data collected using a continuous measure of frequency or duration to the graphs showing percent of intervals obtained by using 10-s MTS or 10-s PIR. Results indicated that MTS was more similar to the duration measures and PIR was more similar to the frequency measures.

Some researchers have extended the conceptualization of error across measurement strategies to not only include the methodological error (i.e., error introduced by the measurement system), but also to include differences in human error engendered by each measurement system (i.e., as each system requires differential response effort and frequency, each system may be differentially associated with data collector errors). Green, McCoy, Burns, and Smith (1982) assessed the impact of both methodological and human error on the accuracy of discontinuous recording methods. In their investigation, they distinguished between two different types of accuracy, within-method and between-methods. Between-methods accuracy, or the total error, is influenced by the error inherent in the discontinuous recording method and error caused by observer mistakes. In contrast, within-method accuracy reflects the amount of errors made by an observer during an observation period. Fifty-four undergraduate participants collected data from a 48-min videotape of a woman twisting her hair. The videotape was broken into 8-min segments in which the total duration of hair twisting varied. Participants were trained via written instructions and were assigned to groups of 18; each group collected data using WIR, PIR, or MTS with 10-s intervals. MTS was more accurate than WIR or PIR in regards to both between-method accuracy and within-method accuracy, but these results were limited to relatively brief interval windows (i.e., 10-s).

Murphy and Harrop (1994) reported similar results when they assessed the amount of human error in MTS and PIR. Sixty trained undergraduate students recorded data using both MTS and PIR. Participants were randomly assigned to one of six groups (three groups were assigned to each measurement procedure that differed in the number of behaviors being simultaneously observed—one, two, or three) and then participants viewed the experimental videotape and collected data. To assess human error, the data collected by the participants in each group were compared to a standard record that was created using the same data collection procedure. Furthermore, the data collected by each participant within a group was compared to each other in order to assess inter-observer agreement. Results showed that regardless of the number of simultaneously recorded behaviors, data collected using MTS was more accurate and had higher inter-observer agreement than PIR. Participants using MTS were also more likely to report being able to collect data for longer observations and with greater efficiently than participants using PIR.

In summary of the available research regarding discontinuous measurement, MTS is characterized by the least amount of human and methodological error relative to PIR and WIR methods for capturing longer duration behaviors (Brulle & Repp, 1984; Harrop & Daniels, 1986; Harrop, Daniels, Foulkes, 1990; Saudagras & Zanolli, 1990), but is less accurate in representing more discrete responses (Meany-Daboul et al., 2007; Repp et al., 1976). More recent research has begun to evaluate boundaries of MTS observation intermittency while retaining measurement accuracy. Hanley, Cammilleri, Tiger, and Ingvarsson (2007) conducted an error analysis of the accuracy of MTS in estimating preschoolers' time allocation to nine simultaneously available activities during free play periods. Initially, data were collected using 5-s MTS intervals during ten, 18-min free play periods. Data were then extracted to estimate the use of longer MTS intervals (i.e., durations of 10, 20, 30, 60, 90, 120, 180, 360, 540, and 1080 s). When data were evaluated on the basis of single activities, duration estimates generated by MTS intervals up to 120 s differed from those of the 5-s MTS by no more than 10% (i.e., marginal error was introduced by extending observation intervals as far as once every 2 min).

Kearns, Edwards, and Tingstrom (1990) found that minimal error was introduced in MTS at even longer duration intervals. These researchers created computer generated streams of behavior occupying 20, 40, 60 and 80% of 60-min sessions. These streams were then scored via MTS with intervals of 30 s, 5 min, 10 min, and 20 min. Results were similar to previous research suggesting that briefer MTS intervals provide more accurate duration estimates than longer MTS intervals, regardless of the level of behavior; however, a majority of the intervals they assessed, up to 20 min, were within 10% of the continuous measure.

Up to this point, researchers have examined the methodological error introduced by intermittent sampling of behavior with the findings that MTS is an accurate discontinuous measurement system when intervals are relatively brief; error increases as interval duration expands. The recommendations based upon this available research would be to keep intervals very short. However, this recommendation fails to account for the increased effort associated with more frequent observation and the extent to which this increased effort may result in increased human error (i.e., error introduced by a human data collector incorrectly coding the presence or absence of a behavior). Human error may increase as a function of (a) the frequency of observations required by the MTS system, (b) the number of simultaneous activities required of the observer, (c) the number of responses being scored at one time, and (d) the actual duration or frequency of the target behavior, (Repp et al., 1976).

Distractions can be minimized or eliminated to minimize their impact upon data collectors in most laboratory-based research, but data are frequently collected in "busy" environments such as classrooms, clinics, and workplaces in applied research. Further, in some instances, data collectors may have multiple simultaneous responsibilities. For instance, teachers may frequently be asked to record the ongoing behavior of a child in their classroom in addition to their numerous other competing responsibilities as an educator (e.g., Bailey, Wolf, & Phillips, 1970; Chafouleas, McDougal, Riley-Tillman, Panahon, & Hilt, 2005; Ellingson, Miltenberger, Stricker, Galensky, & Garlinghouse, 2000; Grey, Honan, McClean, & Daly, 2005; Hay, Nelson, & Hay, 1977, 1980; Kubany & Sloggett, 1973; Lerman, Hovanetz, Strobel, & Tetreault, 2009; Maag & Larsen, 2004; Mizes, Hill, Boone, & Lohr, 1983; Nelson, Hay, Hay, & Carstens, 1977; and Symons, McDonald, & Wehby, 1998). Using frequent intervals (e.g., 5 s) may result in low levels of methodological error as a result of the measurement system, but high levels of human error will likely be introduced relative to less frequent intervals (e.g., 10 min). I propose that error analyses need to consider not only methodological error, but also human error when implemented in more typical settings.

The impact of human error during MTS has not been thoroughly evaluated. To date, there are only two instances in which researchers have investigated the impact of human error on the accuracy of MTS procedures, and have done so with a narrow range of MTS intervals and were conducted in relatively ideal (i.e., non-noisy) environments (i.e., Green et al., 1982; Murphy & Harrop, 1994). Thus, I further evaluated the impact of human error on MTS procedures in a two-part study. The first phase assessed the impact of interval duration upon the human and methodological error introduced into MTS data collected by novice data collectors on the behavior of a student in a video of a simulated classroom with programmed occurrence of target behavior. The second phase replicated the procedures of the first phase, but more closely simulated a typical classroom environment by providing data collectors a simultaneous competing task.

GENERAL METHOD

Participants and Setting

In order to determine the appropriate sample sizes for this investigation, I conducted a power analysis using G*power version 3.0.10 (Faul, Erdfelder, Lang, & Buchner, 2007); a power level of .80 with an alpha level of .05 and a standardized effect size of $f = .25$ was used and suggested a total sample size of 60, or 30 participants per phase. I recruited a total of 72 undergraduate students from the LSU experiment research pool, 36 for each phase, to participate in the current investigation. Participants earned 5 research credits for their participation and I randomly assigned participants to one of three experimental groups across the two phases of the experiment.

I conducted all sessions in one of 3 experimental rooms on the LSU campus. Each room contained at least two desks, two chairs and two computers. The computers were equipped with Microsoft® PowerPoint software and had headphones attached. Each session room could accommodate up to two participants at a time.

Materials

I created three 30-min videos. Each video depicted a simulated classroom environment (i.e., multiple people, academic materials present), and I designated one person in the classroom as the target student who engaged in a repetitive behavior (nail biting). The actors and materials in each classroom video were identical. The only difference between videos was the total time the target student engaged in the repetitive behavior (i.e., approximately 80%, 50%, and 20% of session occupied by the target behavior; termed high, moderate, and low occurrence videos, respectively).

I divided each observation into 600, 3-s intervals in order to script the occurrence and non-occurrence of behavior in these videos. I scripted each interval for the actor to engage or not engage in the target behavior (i.e., for the high occurrence video, the target behavior was scripted to occur in 480 of the 3-s intervals and not occur in 120 of the 3-s intervals; in the moderate occurrence video, 300 of the 3-s intervals were scripted to contain the target behavior and 300 of the 3-s intervals were scripted to not contain the target behavior; and 120 of the 3-s intervals were scripted to contain the target behavior for the low occurrence video while 480 of the 3-s intervals were scripted to not contain the target behavior). I distributed the occurrence and non-occurrence intervals randomly throughout each session. Table 1 below displays descriptive data regarding the bouts of behavior. I then edited the videos in order to eliminate ambient noise and to superimpose tones for each MTS data collection interval (i.e., 1 min, 5 min, 10 min, and 15 min). I also added a chapter mark at the point of each tone.

Table 1: Descriptive Data for the Bouts of Behavior

Video	Average	Min	Max	Total Duration
Low proportion (20%)	4 s	3 s	12 s	360 s
Moderate Proportion (50%)	6 s	3 s	24 s	900 s
High Proportion (80%)	15 s	3 s	78 s	1440 s

I created multiple Microsoft® PowerPoint presentations to present the videos. In each presentation, the same video was presented four times. The only difference across videos in each presentation was the superimposed tone for each MTS data collection interval (i.e., 1 min, 5 min, 10 min, and 15 min). I created the videos this way to allow data to be collected using each MTS data collection interval for each participant. Participants were not told that they were watching the same video multiple times. In order to control for possible carryover effects, I presented the videos in six different sequences across participants (i.e., 1-5-15-10-min; 1-10-15-5-min; 5-15-10-1-min; 5-10-1-15-min; 10-15-1-5-min, 15-5-1-10-min).

I gave additional materials to all the participants, which included an ink pen and a data collection packet (see Appendix A). There were separate data collection sheet for each MTS data collection interval. Each data sheet contained a brief instructional statement at the top of the page and the appropriate number of intervals in which data were to be collected. The interval number was listed above each space in which data was to be marked.

Measurement

Standard Measure. Two trained data collectors (two graduate students with more than 1 year of data collection experience) scored all of the videotapes using continuous duration recording with laptop computers to develop a "gold standard" of accuracy regarding the true occurrence of target behavior during each video. Both observers re-scored each video until they achieved greater than 90% agreement on each tape using proportional agreement within 10-s intervals (95.14% for the low occurrence video, 94.87% for the moderate occurrence video, and 95.40% for the high occurrence video). The primary data collector's record of each session was then used as the "gold standard" by which to determine error types. For each session, I divided the total duration of problem behavior by the total number of time within a session (1800 s) and then converted into a percentage of session occurrence.

Extracted Data. I extracted MTS data from the videos at the exact moment of the tone using the chapter marks that I inserted into each video. For instance, I extracted data for the 60-s MTS by viewing the video on pause at each chapter mark (e.g., 60 s, 120s, 180 s, etc) to determine if the target behavior was occurring at the exact moment of the tone. I then divided the total number of intervals scored as an occurrence by the total number of intervals within each session (30 intervals for 1-min MTS, 6 intervals for 5-min MTS, 3 intervals for 10-min MTS, 2 intervals for 15-min MTS) to obtain a percentage of intervals occurrence score.

Observer Collected Data. Each participant collected data during one 130-min block of time. Participants collected data from the same 30-min videotaped session four times with a brief break between each observation. Each time the participants viewed a video, they used one of the four different MTS intervals (i.e., 1 min, 5 min, 10 min, and 15 min) to collect data. For instance, a participant may collect data for a 30-min video using the 1-min MTS, followed by the 5-min MTS, then the 15-min MTS, and finally the 10-min MTS). Again, I used the same video to control for variability in response characteristics. Following each session, I converted each participant's data into a percentage of intervals occurrence score.

Error Analyses (Dependent Variables)

I defined *Methodological Error* as the error introduced as a result of discontinuous measurement, and I calculated it by comparing the difference in the percent occurrence between the continuous "gold standard" and extracted MTS data. I defined *Human Error* as the error introduced by coding mistakes of the observers and I calculated it by comparing the absolute difference in percentage occurrence obtained from the participants' records to that of the extracted records. Finally, I calculated the *Total Error* by comparing the absolute difference in the percent occurrence from the participants' records to the percent occurrence from "gold standard."

Phase I Method: Controlled Assessment of Error

The purpose of Phase 1 was to assess error introduced during MTS as a function of the MTS interval while controlling for features of the target behavior and minimizing observers' competing responsibilities.

Participants. I recruited 36 participants for this phase of the study with 12 participants per group. The mean age of participants who viewed the low occurrence video was 19.5 years old (range 18 to 21 years old). Only one participant in this group reported having collected data in a similar manner in the past. For the moderate occurrence group, the mean age of the participants was 19.8 (range 18 to 28 years old), and two participants reported collecting data using a similar procedure in the past. Finally, the mean age of participants who viewed the high occurrence video was 20.2 years old (range 18 to 25 years old), and no participants reported collecting data previously using a similar procedure.

Procedure. When participants arrived to the session room, I directed them to a desk, and provided them with an informed consent form and a data collection packet. I then read the following instructions,

> "You will be watching 4 videos. Each video is 30 minutes in duration. Your job is to collect data on the amount of time Libby bites her nails. Libby is in the white shirt with the computer and a pink cup. Nail biting is defined as contact of any of Libby's fingers to her mouth. Whenever you hear a beep, score a "+" if Libby is biting her nails at that exact moment. If she is not biting her nails at that exact moment, mark a "-". Be careful that she is touching her mouth and not her cheek. You are free to look away from the video while not scoring data. However, PLEASE DO NOT USE YOUR CELL PHONE OR ENGAGE IN ANY OTHER ACTIVITY. Again, you are to only record what you see at the moment of the beep into the corresponding box. For each video, there is a separate sheet in your packet and the number of boxes represents the number of beeps for that video. Do not worry if it feels like there hasn't been a beep in awhile or if the beep does not start right away. The only sound in the video is the beeping. When each video is

complete, the screen will turn black and say ‘This video is complete.’ Once the screen blinks, press the space bar. The screen will now say ‘Break Time.’ At this time, you are free to take a brief break. If you do not want to take a break, or when you return from your break, press the space bar to start the next video. Bathrooms and water fountains are on the 2nd floor. Do you have any questions?”

Once I answered any questions from the participant, I instructed the participant to put on their headphones and press the space bar to begin the first video. After each video, the participant had to press the space bar two times before advancing to the next video. Once all four videos were complete, the session was complete and the screen instructed the participant to be sure all forms were complete and to return their data packet to the experimenter. I then debriefed the participant, gave him or her credit for participating, and excused him or her to leave.

Phase 2 Method: Assessment of Error When Distracted

The purpose of Phase 2 was to assess error as a function of MTS interval while systematically introducing competing responsibilities to be more similar to a typical classroom environment.

Participants. I recruited 36 novel participants for this phase of the study with 12 participants per group. The mean age of participants whom viewed the low occurrence video was 19.92 years old (range 18 to 23 years old). None of the participants in this group reported having collected data in a similar manner in the past. For the moderate occurrence group, the mean age of participants was 19.58 (range 18 to 22 years old), and only one participants reported collecting data using a similar procedure in the past. Finally, the mean age of participants who viewed the high occurrence video was 20.5 years old (range 19 to 26 years old), and no participants reported having collected data previously using a similar procedure.

Materials. In addition to the materials mentioned above, I provided participants with a Scantron sheet, instruction sheet regarding how to complete the Scantron, and a packet of "student" papers. The packet was made up of 42 different sheets printed 3 times each and presented in a random order for a total of 126 worksheets. Each page contained two worksheets with 20 problems per page. The worksheets were double-sided and contained 1- and 2-digit addition and subtraction problems with regrouping. These "student" papers had a random number of errors, ranging from 0 to 8 errors per worksheet, with an average of 4 errors per page. The errors consisted of adding when the problem should have been subtracted and vice versa, not properly regrouping, and writing the wrong number (see Appendix F for an example worksheet). Participants were to "grade" each worksheet and to identify the number of "errors" on each sheet and to mark that number on the scantron sheet. Each worksheet was numbered corresponding to the numbers on the Scantron. The instruction packet included instructions for recording (e.g., if the first worksheet had 2 errors, the participant would bubble in "C" for number 1 on the Scantron; if the second worksheet contained 7 errors, the participant would bubble in "A" and "D" for number 7 on the Scantron).

Procedure. The procedures were similar to the previous phase. The only difference was the description of the distractor task in the instructions. The participants were read the following instructions prior to answering any questions and beginning:

> "You are here today to grade some papers and have the chance to earn some money. This is a packet of 126 math worksheets. For each worksheet you grade correctly, you will be entered into a raffle for a $20 giftcard. So, if you grade 50 sheets, you will have 50 chances to win the giftcard. As you grade each worksheet, please use the Scranton to record the number of problems incorrect for each worksheet. The number on the

worksheet corresponds to the number on the Scantron. You will also be watching 4 videos and collecting data on each video. Each video is 30 minutes in duration. Your job is to collect data on the amount of time Libby bites her nails. Libby is in the white shirt with the computer and a pink cup. Nail biting is defined as contact of any of Libby's fingers to her mouth. Whenever you hear a beep, score an "+" if Libby is biting her nails at that exact moment. If she is not biting her nails at that exact moment, mark an "-". Be careful that she is touching her mouth and not her cheek. For each video, there is a separate sheet in your packet and the number of boxes represents the number of beeps for that video. Do not worry if it feels like there hasn't been a beep in awhile or if the beep does not start right away. The only sound in the video is the beeping. Again, you are to only record what you see at the moment of the beep into the corresponding box. You are free to look away from the video while not scoring data. However, PLEASE DO NOT USE YOUR CELL PHONE, TALK TO OTHER PARTICIPANTS, OR ENGAGE IN ANY ADDITIONAL ACTIVITIVES OTHER THAN SCORING THE MATH WORKSHEETS. When each video is complete, the screen will turn black and say 'This video is complete.' Once the screen blinks, press the space bar. The screen will now say 'Break Time.' At this time, you are free to take a brief break. If you do not want to take a break, or when you return from your break, press the space bar to start the next video. Bathrooms and water fountains are on the 2nd floor. Do you have any questions?"

RESULTS

Phase 1: Controlled Assessment of Error

The results of Phase 1 are depicted in Figure 1. Methodological error is displayed in the top panel, human error is depicted in the middle panel, and total error is depicted in the bottom panel.

Methodological Error. The methodological error is the same for both phases of the current investigation since the same videos were used. Again, I calculated the methodological error by comparing the percent duration of problem behavior of data collected using a continuous data collection procedure to data collected using the extracted MTS procedure. When the proportion of problem behavior was low, the amount of error was similar across all MTS intervals, with 1-min intervals having the lowest error at 20.03% and the other three intervals were at 26.70%. When the proportion of problem behavior was moderate, the amount of error decreased and was lowest for the 1-min and 5-min intervals, 1.83% and 5.17%, respectively. The amount of error increased for the 10-min and 15-min intervals to 38.40%. Finally, when the proportion of problem behavior was high, the amount of error remained low for the 1-min and 5-min intervals (each 3.47%), decreased for the 10-min interval (13.20%) and remained high for the 15-min interval (36.80%).

Human Error. As described above, I calculated the human error by comparing the absolute difference in percentage occurrence obtained from the participants' records to the percentage occurrence obtained from the extracted records. When the proportion of problem behavior was low, the amount of error was less than 5% for all MTS intervals, with the 5-min and 10-min intervals having the lowest error at 2.78% and the error for the 1-min interval was 3.89% and the error for the 15-min interval was 4.17%. When the proportion of problem

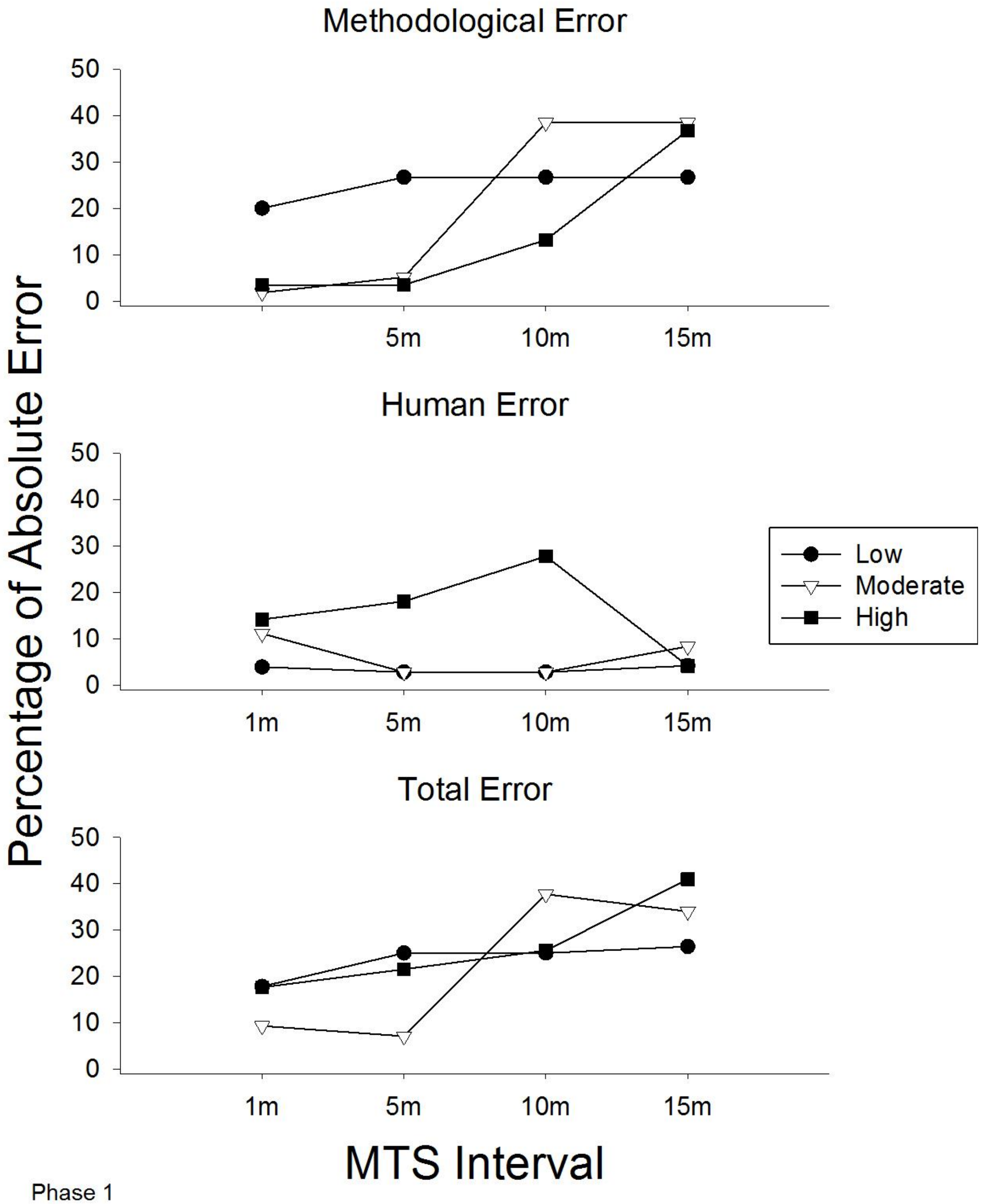

Figure 1. The results of Phase 1. For all panels, MTS interval is along the X-axis and the percent of absolute error is along the Y-axis. The closed squares represent the amount of error for the high percent duration of problem behavior. The open triangles represent the amount of error for the moderate percent duration of problem behavior. The open circles represent the amount of error for the low percent duration of problem behavior.

behavior was moderate, the amount of error increased for the 1-min and 15-min intervals, 11.11% and 8.33%, respectively. The amount of error remained the same, 2.78%, for the 5-min and 10-min intervals. Finally, when the proportion of problem behavior was high, the amount of error increased above 10% for the 1-min, 5-min, and 10-min intervals, 14.17 %, 18.06%, and 27.78 %, respectively. The amount of error for the 15-min intervals was 4.17%, the same as it was when the proportion of problem behavior was low.

I conducted a statistical data analysis, a Mixed Factorial ANOVA, in order to determine if the amount of error for each independent variable was significantly different. To reiterate, the two independent variables consisted of a between-subjects factor and a within-subjects factor. The between-subjects factor was the proportion of problem behavior within a video (3 levels: low, moderate, or high), within-subjects factor was the MTS interval length (4 levels: 1, 5, 10, or 15 minutes), and the dependent variable was percent of human observation error. The interaction between the proportion of problem behavior and MTS interval length was the first effect evaluated, given that a significant interaction effect renders the main effects uninterruptable. A significant interaction suggests that the effect of each independent variable is not independent and thus the two independent variables have an interactive influence of the dependent variable; in other words, one factor depends on the level of another factor. The results are displayed in Table 2 and indicated that there was a significant interaction between the proportion of problem behavior and MTS interval length (F (6) = 5.40, $p < 0.001$). This finding indicated that the effect of the proportion of problem behavior on the human error associated with direct observation depends on the length of the MTS interval. As one can see by examining Figure 1, there appears to be greater differences in human observation error between the interval lengths at certain proportions of problem behavior than others. Given that the degrees of freedom of the interaction effect is greater than 1, indicating a diffuse test, contrast analyses were performed between each of the interval lengths within each level of the proportion of problem behavior.

Table 2: ANOVA Table for Phase 1

Source	*df*	*F*	*p*-value
Proportion of Behavior (between-subjects effect)	2	5.00	0.01
MTS Interval length (within-subjects effect)	3	1.95	0.13
Behavioral variability X Interval length (interaction effect)	6	5.40	< 0.001

The results of the contrast analyses between the MTS interval lengths at each level of the proportion of problem behavior within a video are depicted in Table 3. With regard to low proportion of problem behavior, the contrast analyses revealed no significant differences between the interval lengths. With regard to moderate proportions of problem behavior, the contrast analyses revealed significant differences between the interval lengths. In particular, significant differences were noted for 1 min vs. 5 min and 1 min vs. 10 min. Interpretation of the significant differences based on the mean error for each interval indicated that the 1-min interval was associated with significantly more error than the 5-min and 10-min interval when the proportion of the problem behavior was moderate. With regard to the high proportion of problem behavior, the contrast analyses revealed significant differences between the interval lengths. In particular, significant differences were noted for 1 min vs. 10 min, 5 min vs. 10 min, and 10 min vs. 15 min. Interpretation of the significant differences based on the mean error for each interval indicated that the 1-min, 5-min, and 15-min intervals were associated with significantly less error than the 10-min interval when the proportion of the problem behavior was high.

Total Error. Results are similar to that of the methodological error. When the proportion of problem behavior was low, the amount of error was similar across all MTS intervals, with 1-min intervals having the lowest error at 17.81%, the 5- and 10-min intervals error was each 25.03%, and the 15-min interval's error was 26.42%. When the proportion of problem behavior

Table 3: Pairwise Comparisons for Phase 1

	Behavioral Variability					
	Low Rates		Moderate Rates		High Rates	
Interval Length Contrasts	t^a	*p*-value	t^a	*p*-value	t^a	*p*-value
1 min. vs. 5 min.	1.17	0.27	3.08	0.01*	-1.87	0.09
1 min. vs. 10 min.	1.17	0.27	2.45	0.03*	-2.20	0.05*
1 min. vs. 15 min.	-0.14	0.89	0.49	0.63	1.82	0.10
5 min. vs. 10 min.	0.00	1.00	0.00	1.00	-1.74	0.11
5 min. vs. 10 min.	-1.00	0.34	-1.08	0.31	2.60	0.03*
10 min. vs. 15 min.	-1.00	0.34	-1.30	0.22	2.68	0.02*

[a] $df = 11$
*$p < .05$

was moderate, the amount of error was the lowest for the 1-min and 5-min intervals, 9.38% and 7.00%, respectively. Compared to low proportions of problem behavior, the amount of error increased for the 10-min and 15-min intervals to 35.62% and 33.93%, respectively. Finally, when the proportion of problem behavior was high, the amount of error observed was similar to that when the proportion of problem behavior was low for the 1-min, 5-min, and 10-min intervals (17.63%, 21.52%, 25.58%, respectively) and much higher for the 15-min interval (40.97%). Thus, based upon these results, the differences in error were described predominantly by differences in method error; human error was low and was not affected by the interval.

Phase 2: Assessment of Error When Distracted

Human Error. Again, the first step in the data analysis was to perform a Mixed Factorial ANOVA (see Table 4 for results). The interaction between the proportion of problem behavior and MTS interval length was the first effect evaluated, and the results indicated that there was not a significant interaction between the proportion of problem behavior and MTS interval length (F (6) = 0.84, p = 0.54). Similarly, there was not a significant effect for the proportion of problem behavior (between-subjects effect) or MTS interval length (within-subjects effect).

These finding indicated that the amount of human error was similar across videos with differing proportions of problem behavior and MTS intervals used to score the videos.

Table 4: ANOVA Table for Phase 2

Source	*df*	*F*	*p*-value
Proportion of Behavior (between-subjects effect)	2	2.08	.14
MTS Interval length (within-subjects effect)	3	.43	.74
Behavioral variability X Interval length (interaction effect)	6	.84	.54

Table 5: Pairwise Comparisons for Phase 2

	Behavioral Variability					
	Low Rates		Moderate Rates		High Rates	
Interval Length Contrasts	t^a	*p*-value	t^a	*p*-value	t^a	*p*-value
1 min. vs. 5 min.	0.51	0.62	1.99	0.07**	-0.39	0.71
1 min. vs. 10 min.	2.08	0.06	3.28	0.01*	-0.83	0.43
1 min. vs. 15 min.	-0.50	0.63	0.37	0.72	0.23	0.75
5 min. vs. 10 min.	1.00	0.34	0.32	0.75	-0.29	0.78
5 min. vs. 10 min.	-0.62	0.55	-1.39	0.19	0.48	0.64
10 min. vs. 15 min.	-1.00	0.34	-0.84	0.42	0.64	0.54

[a] *df* = 11
$*p < .05$
** *The p-value is approaching significance of .05*

The results of the contrast analyses between the MTS interval lengths at each level of the proportion of problem behavior within a video are depicted in Table 5. Significant differences were only observed when the proportion of behavior was moderate for the 1 min vs. 5 min and 1 min vs. 10 min, suggesting that the 1-min interval was associated with significantly more error than the 5-min and 10-min intervals when the proportion of the problem behavior was moderate.

As one can see by examining Figure 2, the amount of human error was less than 11% across all proportions of problem behavior and MTS intervals. For the low proportion of problem behavior, human error was under 5% for all MTS intervals and was highest for the 15-min interval at 4.17%. When there was a moderate proportion of problem behavior, there was an increase in the amount of error for all MTS intervals used. Finally, when there was a high proportion of problem behavior, the amount of human error decreased for the 1-min and 15-min MTS interval, but the amount of human error increased again for the 5-min and 10-min MTS intervals (6.94% and 8.33%, respectively).

Total Error. Results are similar to that of the previous phase. When the proportion of problem behavior was low, the amount of error was moderate across all MTS intervals (18.09%, 25.31%, 26.70%, and 26.42 % respectively). When the proportion of problem behavior was moderate, the amount of error decreased for two MTS intervals (i.e., the 1-min and 5-min intervals, 8.82% and 7.54%, respectively) and increased for two of the MTS intervals (i.e., the 10-min and 15-min intervals was 35.62% and 33.93%, respectively). Finally, when the proportion of problem behavior was high, the amount of error continued to increase for the 15-min interval (i.e., 34.83%), decreased for the 10-min interval (i.e., 14.93%), and remained low for the 1-min and 5-min intervals (i.e., 9.30% and 10.41%, respectively).

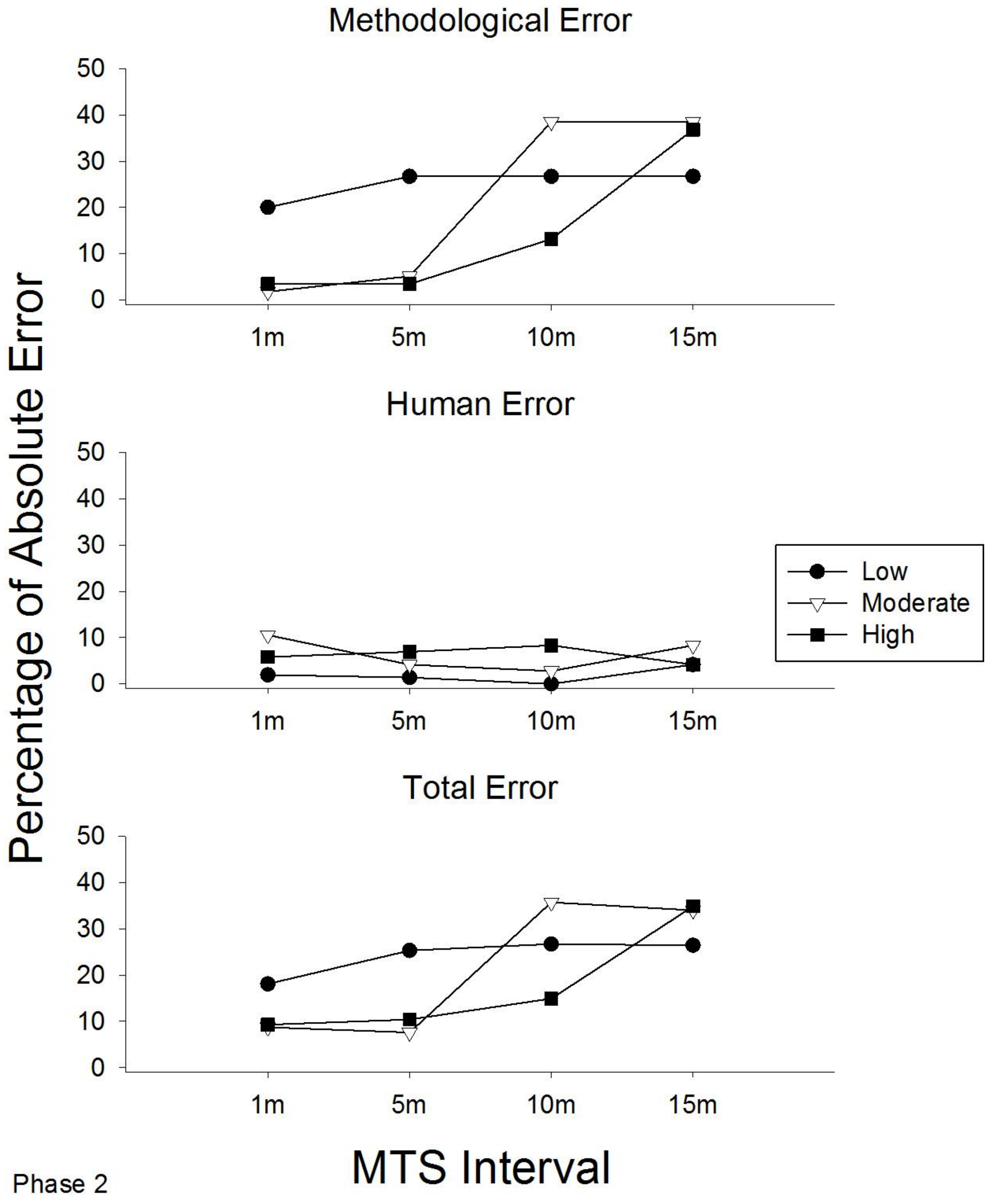

Figure 2. The results of Phase 2. Data are presented in the same way as Figure 1.

DISCUSSION

The purpose of this investigation was to evaluate error introduced during data collection using MTS. I evaluated the amount of methodological, human, and total error introduced during MTS data collection across varied MTS intervals and durations of behavior within a highly controlled experimental context. There were a number of findings from this investigation that I found notable. First, brief intervals resulted in the least amount of methodological error, and error levels increased as observation intervals increased. This finding is consistent with previous research (e.g., Ary, 1984; Gardenier et al., 2004; Harrop & Daniels, 1986; Kearns et al, 1990; Murphy & Goodall, 1979; Powell et al., 1975, 1977; Rapp et al., 2008; and Saudargas & Zanolli, 1990) and extends these evaluations to include interval durations that may be more likely to be used in classroom application. Kearns et al. reported average error rates of less than 10% when using data to simulate 5-, 10- and 20-mintue MTS intervals. In the current study, I also obtained reasonably low levels of methodological error given up to 5-min MTS intervals (with high and moderate duration behaviors, only). However, I calculated unacceptably high error rates for all intervals greater than 5-min.

I also found an interaction between methodological error and behavior duration. Specifically, MTS was more accurate in capturing behavior that occurred at moderate to high durations, but was inaccurate in estimating behavior that occurred at lower duration. Powell et al. (1975) similarly reported that methodological error was less than 20% when MTS intervals were 240 s or less. Similarly, Powell et al. (1977) evaluated the accuracy of data collected using MTS when target behaviors occupied 20%, 50%, and 80% of sessions; these authors reported methodological error of less than 20% for MTS intervals shorter than 120 s. The reason for these findings may be more related to the duration of each individual instance of behavior rather than the total duration across a session. Put simply, brief duration behaviors are less likely to be occurring and therefore captured by the momentary time sampling procedure (Meany-Daboul et al., 2007; Rapp et al., 2008; Repp et al., 1976).

Perhaps the most interesting finding of the current studies was that human error was negligible across all evaluated MTS intervals. This was the case in phase 1 in which I attempted to create an ideal data collection environment and again in phase 2 in which I simulated portions of a classroom by providing participants a distracter task and arranging a reward for engaging in that distracter task. Our assumption prior to conducting this study was that there would be a tradeoff between human and methodological error as the effort of frequent data collection increased with shorter MTS intervals; this was not the case. In fact, human error rates were fairly low across all examined observation frequencies, which speaks to the relative ease of MTS as a data collection system. Instead our results indicate that the accuracy of data is not compromised via human factors and thus, from a methodological perspective, measurement systems in classrooms should be designed with the most frequent data collection possible to limit methodological error.

There are two limiting conditions to this statement that should be noted however. First, the most frequent interval I examined was 1 min, which is fairly infrequent by most behavior-analytic study standards. It may be the case that requiring more frequent data would result in additional human error. Secondly, the accuracy of data will also be impacted by teachers' willingness to collect data in their classroom. Although frequent data collection may not affect the methodological soundness of a measurement system, it may affect the extent to which teachers will agree to participate. An important future area of research will be to assess teacher acceptability of MTS data collection requiring different frequencies of observation and recording.

Hanley et al. (2007) described a procedure similar to the multiple-stimulus without replacement preference assessment procedure described by DeLeon and Iwata (1996) in which they examined data collectors' preference of three different MTS intervals (60 s, 90 s, and 120 s). They presented experienced observers with an array of data sheets each day. Each data sheet represented one of the MTS intervals. After choosing a data sheet, the observers then used that sheet to collect data for the day. The following day, the array again presented before collecting data; however, the chosen data sheet was not included in the array. The resulting rank-order revealed that participants preferred the 90-s interval to the 60-s and 120-s interval. This procedure could be applied to teachers in the classroom once they have had experience with multiple data collection methods. It also may be beneficial to ask teachers which method they prefer once they have had experience since the results of the preference assessment also matched the participants' verbal report about their preferred interval.

In addition to observation frequency, human error may also be impacted by the duration of observation sessions (i.e., fatigue). I did not systematically evaluate the accuracy of data collection across time, but it is worth noting that most measurement studies evaluate relatively brief sessions (e.g., up to 30 min), but teachers would be expected to collect data throughout a class period (60 min) or a full day depending upon the frequency with which they see the target student. Future research should investigate the impact of data collection fatigue as well.

Finally, the extent to which our findings regarding human error can be generalized to classroom environments is unclear. I introduced a small portion of a teacher's responsibilities by requesting our participants grade papers; however actual teachers clearly have more complex duties (e.g., instruction). In particular, grading simple arithmetic is a very discrete event which can be easily interrupted and resumed for data collection whereas a social interaction during instruction may be more difficult to interrupt and therefore be more likely to compete with data collection and introduce error. The current study was a first step in evaluating human error introduced into data collection; additional research to identify factors contributing to human error is warranted.

REFERENCES

Ary, D. (1984). Mathematical explanation of error in duration recording using partial interval, whole interval, and momentary time sampling. *Behavioral Assessment, 6,* 221-228.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*, 91-97.

Bailey, J., Wolf, M., & Phillips, E. (1970). Home-based reinforcement and the modification of predelinquents classroom behavior. *Journal of Applied Behavior Analysis, 3,* 223–233.

Broden, M., Bruce, C., Mitchell, M. A., Carter, V., & Hall. R. V. (1970). Effects of teacher attention of attending behavior of two boys at adjacent desks. *Journal of Applied Behavior Analysis, 3*, 205-211.

Brulle, 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,* 481-485.

Chafouleas, S. M., McDougal, J. L., Riley-Tillman, T. C., Panahon, C. J., & Hilt, A. M. (2005). What do behavior report cards (DBRCs) measure? An initial comparison of DBRCs with direct observation for off-task behavior. *Psychology in the Schools, 42*, 669-676.

Ellingson, S. A., Miltenberger, R. G., Stricker, J., Galensky, T. L., & Garlinghouse, M. (2000). Functional assessment and intervention for challenging behaviors in the classroom by general classroom teachers. *Journal of Positive Behavior Interventions, 2,* 85-97.

Faul, F. (2006). G*power version 3.0.5 [Computer software and manual]. Retrieved from http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/

Gardenier, N. C., MacDonald, R., & Green, G. (2004). Comparison of direct observational methods for measuring stereotypic behavior in children with autism spectrum disorders. *Research in Developmental Disabilities, 25,* 99-118.

Green, S. B., & Alverson, L. G. (1978). A comparison of indirect measures for long-duration behaviors. *Journal of Applied Behavior Analysis*, *11*, 530.

Green, S. B., McCoy, J. F., Burns, K. P., & Smith, A. C. (1982). Accuracy of observational data with whole interval, partial interval, and momentary time-sampling recording techniques. *Journal of Behavioral Assessment, 4,* 103-118.

Grey, I. M., Honan, R., McClean, B., & Daly, M. (2005). Evaluating the effectiveness of teacher training in applied behaviour analysis. *Journal of Intellectual Disabilities*, *9*, 209–227.

Hall, R. V., Lund, D. & Jackson, D. (1968). Effects of teacher attention on study behavior. *Journal of Applied Behavior Analysis*, *1,* 1-12.

Hanley, G. P., Cammilleri, A. P., Tiger, J. H., & Ingvarsson, E. T. (2007). A method for describing preschoolers' activity preferences. *Journal of Applied Behavior Analysis, 40,* 603-618.

Harrop, A. & Daniels, M. (1986). Methods of time sampling: A reappraisal of momentary time sampling and partial interval recording. *Journal of Applied Behavior Analysis, 19,* 73-77.

Harrop, A., Daniels, M., & Foulkes, C. (1990). The use of momentary time sampling and partial interval recording in behavioural research. *Behavioural Psychotherapy, 18,* 121-127.

Hay, L. R., Nelson, R. O., & Hay, W. M. (1977). The use of teachers as behavioral observers. *Journal of Applied Behavior Analysis, 10,* 345-348.

Hay, L. R., Nelson, R. O., & Hay, W. M. (1980). Methodological problems in the use of participant observers. *Journal of Applied Behavior Analysis, 13,* 501-504.

Johnston, J. M, & Pennypacker, H. S. (1980). *Strategies and tactics of behavioral research* (1st ed.). Hillsdale, NJ: Lawrence Erlbaum.

Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Kazdin, A. E., Silverman, N. A., & Sittler, J. L. (1975). The use of prompts to enhance vicarious effects of nonverbal approval. *Journal of Applied Behavior Analysis, 8,* 279-286.

Kearns, K., Edwards, R., & Tingstrom, D. H. (1990). Accuracy of long momentary time-sampling intervals: Implications for classroom data collection. *Journal of Psychoeducational Assessment, 8,* 74-85.

Kubany, E. S., & Sloggett, B. B. Coding procedure for teachers. *Journal of Applied Behavior Analysis, 6,* 339-344.

Lerman, D. C., Hovanetz, A., Strobel, M., & Tetreault, A. (2009). Accuracy of teacher-collected descriptive analysis data: A comparison of narrative and structured recording formats. *Journal of Behavioral Education, 18,* 157-172.

Maag, J. W., & Larsen, P. J. (2004). Training a general education teacher to apply functional assessment. *Education and Treatment of Children, 27,* 26–36.

Mann, J., Have, T. T., Plunkett, J. W., & Meisels, S. J. (1991). Time Sampling: A methodological critique. *Child Development, 62,* 227-241.

Meany-Daboul, M. G., Roscoe, E. M, Bourret, J. C., Ahearn, W. H. (2007). A comparison of momentary time sampling and partial-interval recording for evaluating functional relations. *Journal of Applied Behavior Analysis, 40,* 501-514.

Mizes, J. S., Hill, G. P., Boone, S. E., & Lohr, J. M. (1983). Reactivity of measurement: A single-subject investigation of effects of participant observers. *Psychological Reports, 53,* 663-668.

Murphy, M. J. & Harrop, A. (1994). Observer error in the use of momentary time sampling and partial interval recording. *British Journal of Psychology, 85,* 169-179.

Murphy, G., & Goodall, E. (1980). Measurement error in direct observations: A comparison of common recording methods. *Behaviour Research and Therapy, 18,* 147–150.

Nelson, R. O., Hay, L. R., Hay, W. M., & Carstens, C. B. (1977). The reactivity and accuracy of teachers' self-monitoring of positive and negative classroom verbalizations. *Behavior Therapy, 8,* 972-985.

Powell, J., Martindale, A., & Kulp, S. (1975). An evaluation of time-sample measure of behavior. *Journal of Applied Behavior Analysis, 8,* 463-469.

Powell, J., Martindale, B., Kulp, S., Martindale, A., & Bauman, R. (1977). Taking a closer look: Time sampling and measurement error. *Journal of Applied Behavior Analysis, 10,* 325-332.

Rapp, J. T., Colby-Dirkesen, A. M, Michalski, D. N., Carroll, R. A., & Lindenberg, A. M. (2008). Detecting changes in simulated events using partial-interval recording and momentary time sampling. *Behavioral Interventions, 23,* 237-269.

Repp, A. C., Roberts, D. M., Slack, D. J., Repp, C. F., & Berkler, M. S. (1976). A comparison of frequency, interval, and time sampling methods of data collection. *Journal of Applied Behavior Analysis, 9,* 501-508.

Saudargas, R. A. & Zanolli, K. (1990). Momentary time sampling as an estimate of percentage time: A field validation. *Journal of Applied Behavior Analysis, 23,* 533-537.

Symons, F. J., McDonald, L. M., & Wehby, J. H. (1998). Functional assessment and teacher collected data. *Education and Treatment of Children, 21*, 135-159.

Thomson, C., Holmberg, M, & Baer, D. M. (1974). A brief report on comparison of time-sampling procedures. *Journal of Applied Behavior Analysis, 7,* 623-626.

Volpe, R. J. & McConaughy, S. H. (2005). Systematic direct observational assessment of student behavior: Its use and interpretation in multiple settings: An introduction to the miniseries. *School Psychology Review, 34,* 451-453.

APPENDIX A: DATA COLLECTION PACKET

Participant #:__________ Date:__________

Whenever you hear a beep, score an "+" if Libby is biting her nails at that exact moment. If she is not biting her nails at that exact moment, mark an "-". Nail-biting is defined as contact of any of Libby's fingers to her mouth. Be careful that she is touching her mouth and not her cheek. You are free to look away from the video while not scoring data. You are to only record what you see at the moment of the beep.

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30

Participant #:__________ Date:__________

Whenever you hear a beep, score an "+" if Libby is biting her nails at that exact moment. If she is not biting her nails at that exact moment, mark an "-". Nail-biting is defined as contact of any of Libby's fingers to her mouth. Be careful that she is touching her mouth and not her cheek. You are free to look away from the video while not scoring data. You are to only record what you see at the moment of the beep.

1	2	3	4	5	6

Participant #:__________ Date:__________

Whenever you hear a beep, score an "+" if Libby is biting her nails at that exact moment. If she is not biting her nails at that exact moment, mark an "-". Nail-biting is defined as contact of any of Libby's fingers to her mouth. Be careful that she is touching her mouth and not her cheek. You are free to look away from the video while not scoring data. You are to only record what you see at the moment of the beep.

1	2	3

Participant #:__________ Date:__________

Whenever you hear a beep, score an "+" if Libby is biting her nails at that exact moment. If she is not biting her nails at that exact moment, mark an "-". Nail-biting is defined as contact of any of Libby's fingers to her mouth. Be careful that she is touching her mouth and not her cheek. You are free to look away from the video while not scoring data. You are to only record what you see at the moment of the beep.

1	2

Data Collection Study

Participant #:___________ Date:__________

Age:____________

Major:_________________________

Year in school:______________________________

Have you collected data like this before (please circle one):

Yes No

Comments:

____

____

____

____

APPENDIX B: EXAMPLE MATH WORKSHEET

Curriculum-Based Assessment Mathematics
Multiple-Skills Computation Probe: Student Copy

Student: Date: ____________

78 +574 652	53 -48 101	453 -387 66	565 +49 615
93 -84 9	613 -225 388	678 +286 964	61 -17 44
505 -348 147	338 +74 412	84 -56 140	355 -278 77
29 +88 117	45 -17 32	331 -176 155	67 +76 143
72 -33 39	825 -557 268	76 +35 111	73 -57 16

www.interventioncentral.org

Curriculum-Based Assessment Mathematics
Multiple-Skills Computation Probe: Student Copy

Student: Date: ____________

89 +386 475	74 -25 49	414 -166 248	699 +17 716
42 -27 15	555 -467 88	255 +66 311	32 -24 8
351 -163 178	89 +24 113	82 -36 46	630 -481 149
96 +788 874	33 -24 57	803 -435 368	97 +578 675
72 -43 29	846 -287 559	88 +382 470	95 -37 62

www.interventioncentral.org

VITA

Joslyn Cynkus Mintz is a candidate for a Doctor of Philosophy degree in the school psychology program at Louisiana State University. She graduated with a Bachelor of Science degree from the University of Georgia in 2001. Joslyn received her Master of Science degree in educational psychology from Georgia State University in 2008. While completing this degree, she gained experience working with children with autism under the supervision of Dr. Wayne Fisher, Dr. Cathleen Piazza Dr. Michael Kelley, and Dr. Robert LaRue. At Louisiana State University, she has worked under the supervision of Dr. George Noell and Dr. Jeffery H. Tiger. Joslyn completed her pre-doctoral internship through the Nebraska Internship Consortium in Professional Psychology at the Munroe-Meyer Institute. She has accepted a postdoctoral position through the Emory University School of Medicine Postdoctoral Fellowship Program in Professional Psychology at the Marcus Autism Center in Atlanta, Georgia.