# Measurement of Response Duration and Frequency Using Partial-interval,Whole-interval, and Momentary Time Sampling Procedures

by

David H. C. Chan

A Thesis
Submitted to the Faculty of Graduate Studies
in partial fulfilment of the requirements
for the degree of

Master of Arts

Department of Psychology
University of Manitoba
Winnipeg, Manitoba

(c) December, 1996

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

*Your file Votre référence*

*Our file Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-23245-X

Canada

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

# MEASUREMENT OF RESPONSE DURATION AND FREQUENCY USING PARTIAL-INTERVAL, WHOLE-INTERVAL, AND MOMENTARY TIME SAMPLING PROCEDURES

BY

DAVID H.C. CHAN

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF ARTS

David H.C. Chan 1997 (c)

Permission has been granted to the Library of The University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to Dissertations Abstracts International to publish an abstract of this thesis/practicum.

The author reserves other publication rights, and neither this thesis/practicum nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

## Abstract

Suen and Ary's (1986a) statistical model to specify time sampling intervals that should produce accurate estimates of response duration and frequency was tested. Intervals with low probability of error (appropriate) were compared to those with high probability of error (inappropriate), as determined by the Suen and Ary method. Discrepancy (from real-time data) indices were claculated for partial-interval, whole-interval, and momentary time-sampling methods using data from three social behaviors of two preschool children in 30 min observation sessions. In general, for both duration and frequency measures (for both children), all three time sampling methods produced statistically reliable error in comparison to real-time recording. However, reliably greater error was found with inappropriate than with appropriate intervals, in accord with Suen and Ary's model.

# ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. John Whiteley, and my departmental committee member, Dr. Steve Holborn, for their guidance, encouragement, and assistance throughout this project. I would also like to thank my external examiner, Dr. Michael Mahon for his insightful comments.

I am greatly indebted to Dr. Dell Ducharme, former Clinical Director, and all the team members of the Preschool Early Intervention Program at the Society for Manitobans with Disabilities for their cooperation and assistance during the data collection phase of this project. I am especially grateful to the participants in this project and their parents for their interest and cooperation.

Finally, I wish to thank my wife, Melinda, for her continuous support, understanding, and encouragement.

# TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES vi

LIST OF TABLES vii

Page

INTRODUCTION AND REVIEW OF THE LITERATURE

Observation Procedures 2

Adequacy of Different Sampling Procedures 8

Estimation of Frequency and Duration 15

The Search for Appropriate Interval Length 19

METHOD 34

RESULTS 53

DISCUSSION 62

REFERENCES 71

Appendices

A. Sample of Real Time Data Printout from "Videocount" 82

B. Duration and Frequency Data from Real Time, Partial- interval, Whole-interval, and Momentary Time Sampling Methods and Their Respective Discrepancy Indexes 83

Their Respective Discrepancy Indexes for Ronnie.

C. Duration and Frequency Data from Real Time, Partial-interval, Whole-interval, and Momentary Time Sampling Methods and Their Respective Discrepancy Indexes for Celine. 89

D. Summary of ANOVA with Repeated Measures on Duration and Frequency Estimates from Sampling Methods. 95

E. Extra Complication with Ronnie's data: Two Factor Interaction Between Sampling Methods and Social Behavior, and Between Sampling Interval and Type of Social Behavior. 97

# LIST OF FIGURES


| | | Page |
|---|---|---|
| 1. | Duration estimates as a function of appropriateness of sampling intervals and sampling methods. | 56 |
| 2. | Duration estimates as a function of sampling intervals and sampling methods, using absolute discrepancy indices. | 58 |
| 3. | Frequency estimates as a function of appropriateness of sampling intervals and sampling methods. | 59 |
| 4. | Duration estimates as a function of sampling methods and play behavior. | 98 |
| 5. | Frequency estimates as a function of sampling methods and play behavior. | 99 |
| 6. | Frequency estimates as a function of appropriateness of sampling intervals and play behavior. | 101 |

# LIST OF TABLES


| | | Page |
|---|---|---|
| 1. | Conditions to Minimize Systematic and Random Errors in Sampling Results. | 18 |
| 2. | Computer Keys, Target Behaviors and Operational Definitions. | 37 |
| 3. | Duration of Shortest Behavioral Bout (b) and Shortest Interresponse Time (t) in seconds from Pilot Sessions of Ronnie. | 44 |
| 4. | Duration of Shortest Behavioral Bout (b) and Shortest Interresponse Time (t) in seconds from Pilot Sessions of Celine. | 45 |
| 5. | Probabilities of Having Measurement Errors as a Function of Sampling Interval Appropriateness for Ronnie. | 49 |
| 6. | Probabilities of Having Measurement Errors as a Function of Sampling Interval Appropriateness for Celine. | 50 |
| 7. | Comparisons of Estimates of Duration for Sampling and Real time Measures. | 61 |
| 8. | Comparisons of Estimates of Frequencies for Sampling and Real time Measures. | 63 |

# Measurement of Response Duration and Frequency Using Partial-interval, Whole-interval, and Momentary Time Sampling Procedures

## Introduction and Review of the Literature

Direct observation is a method of data collection in which target behaviors are observed and recorded as they occur by human observers (Repp, Nieminen, Olinger, & Brusca, 1988). Direct observation is generally used in research or clinical settings to describe the behavior of individuals or groups, how their behavior changes over time, or how their behavior covaries with treatment procedures (Kelly, 1977; Repp, 1983). The goal of direct observation is to obtain accurate measures of particular dimensions of target behaviors.

Direct observation of behavior has been a methodological mainstay in many areas of the social sciences and has been widely used in applied research (Foster, Bell-Dolen, & Burge, 1988; Hutt & Hutt, 1970; Wright, 1960). Within the field of psychology, direct observation has its roots in research conducted in preschool settings during the late 1920s and the 1930s (Wasik, 1989; Wright, 1960). From the 1960s onward, direct observation techniques have been used in field

studies, in school classrooms, in laboratory studies of mother-infant interactions, and in experimental as well as applied analyses of behavior (Wasik, 1984, 1989). The wide application of direct observation in various areas of the behavioral sciences warrants examination of its accuracy.

## Observation Procedures

### Real Time Recording

Observation of behavior can be done either by observing every instance of the behavior or by observing selected instances of behavior through different sampling procedures. Recent developments in electronic and computer technology have produced systems that facilitate the recording of behaviors in "real time" (Bass, 1987; Celhoffer, Boukydis, Minde, & Muir, 1977; Fitzpatrick, 1977). In real-time recording, target behaviors are operationally defined and are usually coded for recording purposes. Then target behaviors are observed continuously over the observation session (Repp, 1983). The onset and offset of each instance of behavior can be recorded. For this reason, real-time recording is able to measure the frequency and duration of target behaviors with

relative accuracy in "real time", as they occur (as opposed to estimating frequency or duration when sampling procedures are used).

While real-time recording is capable of yielding both frequency and durational data, and has been accepted as the most accurate ("gold standard") recording procedure (Sanson-Fisher, Poole, Small, & Fleming, 1979), it requires the use of electronic devices in most cases, and can be very costly (in terms of money, time and personpower). Also, real-time recording may place high demands on the observer because the onset and offset of every instance of behavior has to be recorded. The quest for observational procedures which are less costly and less demanding on the observer has sparked research on the feasibility of using sampling procedures in which the observation session is divided into intervals (e.g., Green, McCoy, Burns, & Smith, 1982; Klesges, Woolfrey, & Vollmer, 1985; Powell, Martindale, & Kulp, 1975; Rojahn & Kanoy, 1985; Springer, Brown, & Duncan, 1981). Three main categories of sampling procedures have been

developed: whole-interval sampling, partial-interval sampling, and momentary time sampling (Repp et al., 1988).

Time Sampling Procedures

Time sampling is one of the most frequently utilized techniques in the systematic observation of behavior. Of the studies published in the Journal of Applied Behavior Analysis between 1968 and 1975, 70% (out of a total of 77% that employed direct observation) used time sampling procedures (Kelly, 1977). A review of recent literatures (1990-present) indicated that time sampling is still widely used by researchers, particularly in school settings (e.g., Bay-Hinitz, Peterson, & Quilitch, 1994; Cullen, 1993; Goldstein & Cigar, 1992; Kaiser, Snyder, & Rogers, 1995). Time sampling techniques consist of dividing the observation session into smaller intervals of constant length for which the observer has to make a decision about whether the target behavior has occurred (Quera, 1990). The session may be divided into a sequence of n-second intervals during which observation and recording takes place (e.g., Ary, 1984; Ary & Suen, 1983; Green et al., 1982; Suen & Ary, 1986a), or

alternatively, the session may be divided into a sequence of observation intervals and recording intervals (e.g., Bijou, Peterson, & Ault, 1968; Harrop & Daniels, 1986; Repp, Roberts, Slack, Repp & Berkler, 1976). The best known types of time sampling procedures are partial-interval sampling (or partial sampling), whole-interval sampling (or whole sampling), and momentary time (or instantaneous) sampling (Altmann, 1974; Bakeman & Gottman, 1987; Powell et al., 1975; Suen & Ary, 1989).

For partial-interval (or one-zero) recording, the observer scores "one" if the behavior occurs at any time during the interval, and "zero" if it does not occur at all during the interval. An interval scored "one" may contain a single response, or multiple responses, or the termination (or continuation) of a previously initiated response, or a combination of these alternatives. In other words, intervals are sometimes scored as a function of the frequency characteristics of the behavior, or at times, as a function of the durational characteristics of the behavior (Powell, 1984a). A behavior will only be scored once regardless of its actual number of

occurrences within an interval (Green et al., 1982; Powell et al., 1975). Thus, information may be lost if the response occurs more than once during the interval. For whole-interval recording, the observer scores "one" if the behavior lasts throughout the entire interval and "zero" if it occurs during only part of the interval or never occurs. For momentary sampling, the observer scores "one" if the behavior occurs at a particular instant (usually the beginning or the end of the interval) and "zero" if it does not (Altmann, 1984; Ary & Suen, 1983; Suen & Ary, 1986a; Suen & Ary, 1986b; Powell, 1984a; Powell, 1984b).

The advantages of sampling procedures are that they do not necessarily require any electronic devices which are expensive and which may be obtrusive in the observational setting (Ary & Suen, 1983). Sampling procedures are also less demanding on the observer in that the observer is not required to record onset and offset of each behavioral episode (because a particular behavior will only be scored once during an interval). This is especially true with time sampling procedures in which observations are conducted only at selected intervals (e.g., 10-s observe, 10-s record partial-

interval sampling) or at the end of the interval (e.g., momentary time sampling). Although such procedures may incur further loss of information, they enable the observer to observe more than one participant within a session (sequential time sampling), as opposed to real-time recording in which one observer can only observe one participant throughout a session. For example, consider that a 10 min session is divided into 60 10-s intervals and a partial interval sequential sampling procedure (see Suen & Ary, 1986a) is employed. The observer can observe Participant 1 in the first interval and Participant 2 in the second interval and so on until all participants have been observed and the occurrence or non-occurrence of the target behavior has been recorded. The observer can then observe Participant 1 again and repeat the process until the end of the session.

The disadvantage of sampling procedures is that important information may be lost, because not all instances of behavior are being recorded. Such loss of information may affect data accuracy. A major concern, then, is whether the aforementioned sampling procedures yield similarly accurate data to those yielded by real-

time recording (see Lipinski & Nelson, 1974; Repp et al., 1988). If they do not differ significantly as determined by statistical analysis, or if the discrepancies, as reflected in the form of a discrepancy index, are acceptable to the researcher, then the savings in time and labour will justify the use of sampling procedures in lieu of the more costly real-time recording.

## Adequacy of Different Sampling Procedures

Research has shown that whole- and partial-interval and momentary time sampling differ from one another in terms of accuracy (Foster et al., 1988; Green et al., 1982; Powell, 1984a; Repp et al., 1976). Partial-interval sampling was found to systematically overestimate the duration of behavior (Harrop & Daniels, 1986; Green & Alverson, 1978; Powell, Martindale, Kulp, Martindale, & Bauman, 1977) because "the proportion of scored intervals does not increase linearly as the duration of the behavior increases. The proportion of scored intervals reaches unity before the true duration reaches 100%" (Ary, 1984, p. 222). To illustrate, consider a high frequency, short duration behavior, if the behavior occurs during part,

but not all, of the interval in every interval, it will be scored "one" in every interval, resulting in the proportion of scored intervals being 100% before the true duration reaches 100%.

Whole-interval sampling was found to systematically underestimate the duration of behavior (Ary & Suen, 1983; Powell et al., 1977) because "the proportion of scored intervals does not increase linearly as the duration of the behavior increases. The proportion of scored intervals reaches zero before the true duration reaches 0%" (Ary, 1984, p.222). Using the same high frequency, short duration behavior mentioned above as an example, under whole-interval recording, all intervals will be scored "zero", resulting in the proportion of scored intervals being zero while the true duration is not zero.

Momentary time sampling was found to yield the most accurate estimate of duration (Altmann, 1974; Ary & Suen, 1983; Bruille & Repp, 1984; Repp, Barton, & Brulle, 1982; Saudargas & Lentz, 1986; Saudargas & Zanolli, 1990; Slate & Saudargas, 1986) because its sampling errors are random and do not systematically over- or underestimate the extent of the target

behavior (Ary, 1984). Momentary time sampling is as likely to overestimate as to underestimate behavior duration, and the absolute value of measurement errors tends to be lower than with the other two sampling procedures (Powell et al., 1977). In addition, Rhine and Linville (1980) found momentary time sampling outcomes to be more highly correlated with actual duration data ($r$ = .99 to 1.00) than were partial-interval outcomes ($r$ = .72 to .95). These outcomes can be explained logically.

The logical explanation, as offered by Ary and Suen (1983), is that if the target behavior occurs throughout the entire interval, it will be scored "one" with any of the three sampling procedures. If the behavior does not occur at all, it will be scored "zero" with the three sampling procedures. The behavior will be scored differently under different sampling procedures, only if there is a change of behavioral state (from non-occurrence to occurrence or vice vera) during the interval. If the behavior occurs during part of the interval (called a mixed interval), the interval will be scored "one" under partial-interval recording and "zero" under whole-interval

recording. With momentary time sampling, mixed intervals will sometimes be scored "one" and sometimes be scored "zero", depending on where the target behavior occurs during the interval. Consequently, with partial-interval sampling, all mixed intervals will be added to the intervals during which the behavior occurred throughout the entire interval, producing a maximum overestimate of the true duration of the behavior. With whole-interval recording, only those intervals in which the behavior occurs throughout the entire interval will be scored and no mixed intervals will be scored, producing a maximum underestimate of the true duration (due to the loss of information). With momentary time sampling, some of the mixed intervals (those in which the behavior occurs at the first or the last instant) may be added to the intervals in which the behavior occurs throughout the entire interval. Those mixed intervals in which the behavior does not occur at the first (or the last) instant will not be added. Thus there is no systematic over- or under-estimation of the duration of the behavior. Statistically, the expected mean of duration error with momentary time sampling is zero when the

standard sampling theory is applied (Cochran, 1977). Hence, momentary time sampling yields a better estimate of true duration than whole- and partial-interval sampling. The drawback of momentary time sampling is that it is not sensitive to behaviors that occur for a very low percentage of time (Mansell, 1985; Saudargas & Zanolli, 1990).

To summarize, If a researcher is interested in duration, and decides to employ a sampling method, momentary time sampling would be the method of choice because it yields the most accurate estimates of duration. Whole-interval sampling produces maximum underestimation of duration. Partial-interval sampling produces maximum overestimation of duration. If a researcher is interested in frequency only, a count of frequency of initiation of the response would be the method of choice (Ary & Suen, 1983).

Whether partial-interval sampling constitutes an adequate measurement technique has long been investigated. Researchers such as Altmann (1974), Kraemer (1979), and Powell (1984a) criticized partial-interval sampling as a measurement technique. These authors commented that because partial-interval

sampling results exist as some composite function of both frequency and durational characteristics of the behavior, and because the frequency and durational properties of the behavior are submerged into a single entity (the percent of intervals scored), the composition of any change in the percent of intervals scored is unknown (Powell, 1984a). For example, a change in the intervals scored "one" between baseline and treatment phase could be due to a change in duration, a change in frequency, or a change in both frequency and duration. To illustrate, consider the case where a child with attention deficit disorder is under treatment to increase "on-task" behavior. Suppose the percent of intervals scored "one" is 20 during baseline, and after treatment, the percent of intervals scored "one" has increased to 60. This increase in intervals scored "one" could be a result of increased duration of "on-task" behavior (treatment success), or a decrease in time on-task and an increase in frequency in switching "on and off" the task (treatment failure). For this reason, Powell (1984a) described partial-interval sampling as an "inadequate measurement technique" because "it conceals the

fundamental nature of the phenomena it addresses, and produces an inconsistent result" (p. 218). Altmann (1974) and Kraemer (1979) even suggested the abandonment of partial-interval sampling. On the other hand, Ary and Suen (1983) suggested that because the resultant scores of partial-interval sampling are a function of both duration and frequency, partial-interval sampling would be the method of choice, if you wish the sampling results to reflect a combination of both duration and frequency. Altmann (1974), however, saw this characteristic of partial-interval scores as a weakness rather than as an advantage, as this method does not provide accurate information about either duration or frequency. Thus, whether partial-interval sampling constitutes an adequate measurement technique warrants further investigation.

Although momentary time sampling had been found to produce the most accurate durational estimates among the three sampling procedures, it may not be feasible for the observation of certain behaviors (e.g.,social behaviors with complex coding rules), if the interval is too short (e.g., 5 s or less). Momentary time sampling requires the observer to briefly scan the

situation at specified moments to code the behavior. Given the complexity involved in the coding of social behaviors, it is very difficult for observers to determine the occurrence or non-occurrence of such target behaviors with a brief visual scan. Under these circumstances, partial-interval sampling would be a reasonable alternative; adding to the urgency of examining whether partial-interval sampling is an adequate measurement technique.

## Estimation of Frequency and Duration

With sampling procedures, behavioral dimensions such as frequency and duration can only be estimated from the data collected on the behavior sampled. Green and Alverson (1978), and Green et al. (1982) found that the accuracy of duration estimates is a function of the length of the observation interval in relation to the duration of the target behavior (bout length) and of the interresponse time (IRT, time between the offset and onset of behavior). Suen and Ary (1984) specifically suggested that in order to obtain unbiased results for duration and frequency estimates, the observation interval length must be: (a) shorter than the shortest behavioral bout and shorter than one-half

of the shortest interresponse time for partial-interval recording, (b) shorter than the shortest behavioral bout and shortest interresponse time for momentary time sampling, and (c) shorter than one-half of the shortest behavioral bout and shorter than the shortest interresponse time for whole-interval recording. A summary of these three conditions is presented in Table 1. These three conditions, which reflect the relationship between interval length, shortest behavioral bout, and shortest interresponse time, are based on theoretical explanation of previous research findings (Suen & Ary, 1984). The interval length selected by researchers can, therefore, be of critical importance in determining accuracy of obtained data (Sanson-Fisher, Poole, & Dunn, 1980).

To eliminate systematic errors inherent in the sampling procedures and to minimize random errors, Ary (1984), Ary and Suen (1983), Suen and Ary (1986a, b) developed a quantitative method to estimate frequency and duration from sampled data. These authors suggested that if the observation session is divided into smaller intervals of constant length and the behavior is scored as either "one" (occurrence) or

Table 1

Conditions to Minimize Systematic and Random Errors in Sampling Results

| Sampling Procedures | Conditions |
|---|---|
| Partial-interval | i < b and i < .5t |
| Whole-interval | i < .5b and i < t |
| Momentary time sampling | i < b and i < t |

i = interval length

b = shortest behavioral bout

t = shortest interresponse time

"zero" (non-occurrence) according to the rules of sampling methods, we would obtain a chain of "zero"s and "one"s. Any changes in the state of behavior from occurring to non-occurring or vice-versa) will affect the way the intervals are being scored. For example, a behavior not displayed in the first interval but displayed during the second interval would result in a score of 0 in the first interval and a score of 1 in the second interval.

Frequency is estimated by counting the number of intervals scored "one" that are immediately preceded by a "zero". In addition, if the first interval of the chain is scored 1, the frequency count should be adjusted by +1 (Suen & Ary, 1986). For example, consider a 60-s session which is divided into six 10-s intervals. If the behavior occurs throughout the entire 60 s, and whole-interval (or partial-interval) recording is used, each of the six intervals will be scored as 1. As none of these "one"s are preceded by a "0", the frequency count would be zero. However, because the first interval of the chain is scored 1, the frequency estimate should be adjusted by +1, thus, yielding a frequency estimate of one, which is exactly

the same as the actual frequency of occurrence.

Duration is defined as the proportion of time during the entire session in which the behavior of interest occurs (Ary & Suen, 1983; Suen & Ary, 1986a). The duration estimate is represented by the ratio between the number of intervals scored and the total number of intervals in an observation session (Ary & Suen, 1983; Suen & Ary, 1986a; Quera, 1990). Using the above example for the frequency estimate, the duration estimate would be 6/6 which is 100%. Because the behavior occurs throughout the entire 60 s session, it does occur 100% of the time in which observation is taking place. However, Suen and Ary (1986a) cautioned that these estimates are accurate (unbiased) only if the conditions listed in Table 1 are satisfied. Thus the interval length selected by the researchers is of critical importance to the accuracy of the frequency and duration estimates.

## The Search for Appropriate Interval Length

Despite research evidence suggesting that data accuracy is a function of the interval length in relation to the duration of behavior as well as to interresponse time (e.g., Green et al., 1982; Ary &

Suen, 1983), the determination of observational interval length is still largely an arbitrary matter rather than an empirical one (Sanson-Fisher, Poole, & Dunn, 1980). However, a few researchers (e.g., Mansell, 1985; Sanson-Fisher et al., 1980) did attempt to find ways to determine the appropriate interval length that would allow for accurate data collection. Currently, two methods have been developed to determine the appropriateness of observational interval length: the empirical approach as proposed by Sanson-Fisher et al., (1980), and the statistical approach as suggested by Suen & Ary (1986a).

Empirical Approach

This approach involves using real time recording as the "gold" standard. Data collected under sampling procedures with various interval lengths would then be compared with those from real time recording (Powell, 1984a). The interval length that yields the most accurate data (i.e., data closest to those from real time recording) would be the appropriate interval length. For example, Sanson-Fisher et al. (1980) suggested an empirical method to determine the appropriate interval length. These authors employed

real time recording as the standard for comparison with the data obtained from intervals ranging from 2, 3, 4, 5, 6, to 10 s. The behaviors (e.g., on task, talk) of inpatients in an acute short-stay psychiatric unit were observed. These authors found that a 3-s interval length was capable of capturing 79.6% of all target behavior recorded under real time. Based on these data, the authors suggested that a 3-s interval length might be acceptable for recording both the frequency and durational data given the target behaviors and the participant group. In general, current literature (e.g., Edwards, Kearns, & Tingstrom, 1991; Repp et al., 1982) seems to suggest that observation intervals under 20 s yield accurate data.

The drawback of such an empirical approach is the generalizability of the optimal interval length to other conditions. Sanson-Fisher et al. (1980) cautioned that different participants, settings, or categories of behaviors may require different optimal interval lengths. An interval length that is appropriate for one behavior (or setting) may not be appropriate for another behavior (or setting). Hence, whether the empirically selected "optimal" interval

length is still appropriate under different conditions is questionable. Another drawback of the empirical approach is that it is laborious and tedious to conduct (Rojahn & Kanoy, 1985). For example, the method suggested by Sanson-Fisher et al. (1980) involved comparing data from real time recording with data from each of the (2, 3, 4, 5, 6, and 10 s) momentary time sampling procedures in order to identify an interval length that yields accurate data. Such labours may have deterred researchers from adopting the practice and made the methodology impractical. As data accuracy is a function of the interval length selected, a more efficient method to determine appropriate interval length would greatly facilitate research.

Statistical Approach

The statistical approach, as proposed by Suen and Ary (1986a), provides a procedure that enables researchers to verify the appropriateness of the selected interval length for whole-interval, partial-

interval and momentary time sampling procedures a priori. The statistical approach (Suen & Ary, 1986a) is more efficient in that it enables researchers to calculate the probabilities of having measurement errors in the sampling results for any chosen interval length. The statistical approach can be applied to a single participant or multiple participants (sequential time sampling) observations. This approach also allows for post hoc correction of measurement errors in time sampling duration estimates, if the interval length was arbitrarily chosen and the probability of error assessment had not been done a priori (Suen & Ary, 1986b). In addition, Quera (1990) offers a procedure to calculate measurement errors in sampling procedures when the observation session is divided into a number of equal interval observation and recording periods.

To illustrate Suen and Ary's procedure, if we collect pilot data and obtain the values of both the shortest behavior bout (b) and the shortest interresponse time (t) of the target behavior, we can, based on the values of b and t, choose an interval length (i) so that the relationship between i, b, and t will satisfy the conditions for each sampling methods

listed in Table 1. Such an interval length will lead to unbiased results from the sampling procedures (Suen & Ary, 1986a, 1986b).

The question remains then, as to whether the selected interval length still will be appropriate in future observation sessions. In other words, we need to find out how likely the selected interval length will still be shorter than $\underline{b}$, $.5\underline{b}$, $\underline{t}$, or $.5\underline{t}$, as appropriate, in future sessions.. This can be done by using the Poisson distribution as the underlying distribution of $\underline{b}$ and $\underline{t}$ respectively. Then, by using the Poisson density function, the probability of whether the chosen interval length will violate the conditions stated in Table 1 in future sessions can be calculated. If the probability is low, the probability of measurement errors existing in the sampling results is low. Such a probability is taken as an indication that the sampling results are likely unbiased. On the other hand, if the probability of these conditions being violated is high, the duration or frequency estimates will very likely be biased. Thus, by finding the probability of violating the conditions listed in Table 1, we can examine the appropriateness of the

selected interval length. The specific procedures are explained below.

Suen and Ary (1986b) suggested that the values of the shortest behavioral bout (b) and shortest interresponse time (t) may be assessed from data of pilot sessions using real time recording. The researcher can then choose an interval length so that the relationship between the interval length, shortest behavioral bout, and shortest interresponse time will satisfy the three conditions listed in Table 1. However, because the durations of behaviors sampled may change from one session to another, the length of the shortest behavioral bout (b) and interresponse time (t) may vary from session to session. Thus interval length chosen for a target behavior based on knowledge of the behavior from previous observation sessions may not apply to future observation sessions.

To overcome this problem, Suen and Ary (1986a) proposed that if the values of shortest behavioral bout (b) and shortest interresponse time (t) change from one observation session to another, their values across a number of observation sessions will form a distribution of b and a distribution of t values. These authors

further suggested that the respective distribution of $\underline{b}$ and $\underline{t}$ will approximate a Poisson distribution. The reason is that $\underline{b}$ and $\underline{t}$ can be treated as two continuous random variables which may take on any positive values (i.e., $0 \leq \underline{b}, \underline{t} \leq +\infty$). Among K independent observation sessions, the distribution of the K values of $\underline{b}$ and $\underline{t}$ are extremely positively skewed with the majority of the values clustered near the zero end of the continuum (e.g., it would be extremely unlikely to find a $\underline{b}$ or $\underline{t}$ of 20,000 s; on the other hand, it is impossible to find a $\underline{b}$ or $\underline{t}$ of less than 0 s). If we view the random variable $\underline{b}$ as the amount of time elapsed before behavior termination and the variable $\underline{t}$ as the amount of time elapsed before IRT termination, both distributions for $\underline{b}$ and $\underline{t}$ will closely resemble a Gamma distribution of a random variable $\underline{x}$ which also has the properties of $0 \leq X \leq +\infty$, and of positive skewness; and which is used to assess the probability of X units of elapsed time prior to success (behavior or IRT termination; Hays, 1973). Hence the distributions of $\underline{b}$ and $\underline{t}$ can be treated as two Gamma distributions. Because interval length in sampling procedures is generally set at discrete values (e.g., 3 s), the

Poisson distribution, which is equivalent to the Gamma distribution, for discrete variables, can be used to approximate the Gamma distribution to avoid the tedious task of computing the Gamma function (Suen & Ary, 1986a).

Values of $\underline{b}$ and $\underline{t}$ obtained from pilot sessions can be used to estimate the parameters (mean values) of the $\underline{b}$ and $\underline{t}$ distributions. Once the values of the shortest behavioral bout and the shortest interresponse time are known, an interval length ($\underline{i}$) can then be selected such that $\underline{i}$ is less than the shortest behavioral bout ($\underline{b}$) and one half of the shortest interresponse time ($\underline{t}$) among the pilot sessions, for partial-interval sampling, is less than one half of the shortest behavioral bout ($\underline{b}$) and the shortest interresponse time ($\underline{t}$) among the pilot sessions, for whole-interval sampling, and is less than $\underline{b}$ and $\underline{t}$, for momentary time sampling. From these values of $\underline{b}$ and $\underline{t}$, the means of $\underline{b}$ and of $\underline{t}$ ($\underline{M}b$ and $\underline{M}t$), which are the process intensities for the two Poisson distributions, can also be calculated. The values of $\underline{b}$ and $\underline{t}$ will form a distribution of $\underline{b}$ and a distribution of $\underline{t}$ respectively. As the distribution of $\underline{b}$ and the distribution of $\underline{t}$ is

expected to approximate the Poisson distribution (Suen & Ary, 1986a), the Poisson cumulative density function can be used to assess the probability that the selected interval length would be longer than b, .5b, t, or .5t in future observation sessions. The equations to determine probabilities for the distributions of b and t respectively are as follows:

$$p(x \leq i) = \sum_{j=0}^{i} (e^{-B} B^j / j!) \qquad (1)$$

and

$$p(y \leq i) = \sum_{j=0}^{i} (e^{-T} T^j / j!) \qquad (2)$$

where

x = b* for partial-interval sampling,

.5b* for whole-interval sampling, and

b* for momentary time sampling.

y = .5t* for partial-interval sampling,

t* for whole-interval sampling, and

$t^*$ for momentary time sampling.

e = 2.718 (universal mathematical constant)

i = the selected interval length.

$j$ = 0, 1, 2, 3, ...$i$.

B = $M$b for partial-interval sampling,

.5$M$b for whole-interval sampling, and

$M$b for momentary time sampling.

T = .5$M$t for partial-interval sampling,

$M$t for whole-interval sampling,

$M$t for momentary time sampling.

As the shortest behavioral bout ($b$) and the shortest interresponse time ($t$) are mutually independent, the total probability that the conditions listed in Table 1 are violated is additive and is the sum of the two cumulative probabilities from equations 1 and 2. Hence, the equations for calculating the probabilities of systematic errors for partial-, and whole-interval sampling procedures are presented as follows:

For partial-interval sampling:

$$p(\textit{systematic}\ \epsilon) = p(b^* \leq i) + p(.5t^* \leq i) \qquad (3)$$

For whole-interval sampling:

$$p(\textit{systematic e}) = p(.5\underline{b}^* \leq i) + p(\underline{t}^* \leq i) \qquad (4)$$

For momentary time sampling:

$$p(\textit{systematic e}) = p(\underline{b}^* \leq i) + p(\underline{t}^* \leq i) \qquad (5)$$

where

$\underline{b}^*$ = unknown value of $\underline{b}$ in future observation session.

$\underline{t}^*$ = unknown value of $\underline{t}$ in future observation session.

Although the values of $\underline{b}^*$ and $\underline{t}^*$ are unknown, the probabilities on the right hand side of equation 3 and 4 can be calculated by using the $\underline{b}$ and $\underline{t}$ distributions with $\underline{Mb}$ and $\underline{Mt}$ (from pilot sessions) as the unbiased parameter estimates. Each probability represents the probability value that the conditions associated with the specific sampling procedures (see Table 1) have been violated; and hence, it would indicate the probability that measurement errors exist in the respective time sampling results, if the selected interval length is used. If the probability

of violation is low, the frequency and duration estimate will likely be unbiased. On the other hand, if the probability of these conditions being violated is high, the duration or frequency estimates very likely will be biased.

The statistical approach, though promising, has been rarely used in current research (Suen & Ary, 1986a). The determination of interval length is still largely arbitrary (Sanson-Fisher et al., 1980). Although Suen and Ary (1986a) have provided a detailed theoretical explanation and numerical illustration of the statistical model in the determination of interval length, no field study appears to have been conducted to test the validity of the model. My research examined whether the procedures proposed by Suen and

Ary (1986a) are applicable in a field setting involving observation of children's play. Specifically, the quantitative method of assessing the appropriateness of interval length as suggested by Suen and Ary (1986a) was applied. Intervals with low probability of having measurement error were categorized as appropriate while those with high probability of having measurement error were categorized as inappropriate. The basis of Suen and Ary's model (1986a) lies in determining the probability of variation in shortest behavioral bout and shortest interresponse time; consequently, behaviors that fluctuate in duration and frequency from session to session would be most suitable for testing the model. Social behaviors of preschool children, such as sharing, solitary play, and parallel play satisfy this requirement because they vary in duration and frequency from one session to another (Chan, 1992). In addition, the display of social behaviors between children with disabilities and their non-disabled peers is one of the most important outcomes measured in evaluations of school integration programs (Guralnick, 1990; Haring & Breen, 1989, 1992; Hundert & Hopkins, 1992; Struffy & Struffy, 1985).

Observational procedures that yield accurate data on social behaviors are instrumental to the success of such evaluations. Hence, social behaviors (sharing and cooperating, solitary play, and parallel play) of preschool children were observed.

The hypothesis was that if interval length were selected on the basis of the pilot data using the method of Suen and Ary (1986a), the probability of having measurement errors in partial-interval, whole-interval, and momentary time sampling could be minimized. Sampling intervals with low probability of having measurement errors (appropriate intervals) would likely yield unbiased results that closely approximate real time data. Likewise, sampling intervals with high probability of having measurement errors (inappropriate intervals) would likely yield biased results. The observations using real time, partial-interval, whole-interval, and momentary time sampling procedures were replicated across three social behaviors and two participants.

## Method

The videotapes of children's social behavior used in this study were obtained in an earlier study by Chan (1992). The participants, settings, behavioral code, equipment, observer training, and procedures used by Chan (1992) will be described. The procedures used in the present study for analyzing this set of observations will then be presented.

### Participants

The participants were two 5-year-old children (one male, one female) attending an early intervention program at the Society for Manitobans with Disabilities in Winnipeg, Manitoba, Canada. They had been diagnosed as having a speech disorder, particularly in speech articulation; otherwise they enjoyed good physical and mental health. Selection of participants was based on teachers' recommendations according to the following criteria: (a) the participants were developing at an age appropriate level, permitting age-appropriate cognitive and motor activities; (b) the participants were interacting well with their peers as well as with their teachers; (c) the participants had no apparent

behavioral disorder. Written parental consent was obtained prior to the research.

Playroom Setting

Play activities were observed during a 30-min free play session in a classroom (18 m x 18 m) at the Society for Manitobans with Disabilities. The free play session was scheduled from 1 to 1:30 p.m. daily and was a component of the early intervention preschool program. The program operated four days per week (Monday, Tuesday, Thursday, and Friday) and was composed of morning and afternoon sessions. The staff complement included two preschool teachers, one speech and language therapist, one occupational therapist, and one physical therapist.

The classroom was an open area with activity centres, such as computer centre, house centre, language centre, and arts and crafts centre. Preschool toys, such as dolls, were also available. The free play time was generally child-directed and allowed for ample social interaction and display of social skills. The participants' peer group consisted of 11 children of similar age ($M$ = 5 years) with mild language

disabilities. All children (5 females and 6 males) communicated verbally and non-verbally.

Target Behaviors and Coding

The three target behaviors were sharing and cooperating, solitary play, and parallel play. These behaviors were defined according to a modification of the Parten (1932) scale and other recent research (Hundert & Hopkins, 1992; Lyytinen, 1991). Target behaviors were represented by the numbers 2, 4, and 5 respectively on the computer keyboard for recording purposes (see Table 2). Non-occurrence of targeted behavior was represented by the number zero.

Equipment

A Canon E440 8mm video camera and recorder with a remote omni-directional microphone (Realistic, Model FX101) were used to record the participants' behaviors across all phases of the study. A computer program called "Videocount" specially developed for the research was used to record the frequencies and durations of observed behaviors in real time recording. Data input was accomplished by pressing specific numbers on the computer keyboard that corresponded with

Table 2

Computer Keys, Target Behaviors and Operational Definitions

| Key | Behavior | Operational Definition |
|---|---|---|
| 0 | Null behavior | a) Engaging in behavior that is not defined.<br>b) Behavior that occurs out of view of the camera.<br>c) On looker behavior, the child is a passive observer of classroom activities.<br>d) Period of transition when the child is moving from one space to another. |
| 2 | Sharing/ cooperating | Sharing /cooperating play can involve vocal and/or motor gestural interaction, and can occur under one of two conditions: |

| | | |
|---|---|---|
| Table 2 cont | | i) the participant offers to share or trade materials with a peer, (e.g., offering a crayon).<br>ii) the participant simultaneously uses the same material as his/her peer in working towards a common goal (e.g., coloring on the same piece of paper), or is turn-taking (e.g., playing a game of cards). |
| 4 | Solitary Play | Child plays by him/herself in manipulation of an object which is either appropriate to the object (e.g., throwing/ kicking a ball, not sitting on it), or inappropriate to the object (e.g., using the truck as an airplane), but where the child is |

| | | |
|---|---|---|
| Table 2 cont | | visually attending to the object. |
| 5 | Parallel Play | Children playing alongside one another, using the same material and/or engaging in the same activity but not working towards a common goal (e.g., a group of children independently building sand castles in the sand table; playing independently in the water table; or sitting together at a table independently performing the same tasks - coloring, puzzles, or arts/crafts activity). Parallel play also includes the independent play of children within a particular theme area (e.g., children playing in a "house" or |

| | |
|---|---|
| Table 2 cont | "shopping" play centre, but not engaging in cooperative play; or children sitting at a table independently performing the same type of activity (puzzles, coloring, lego, etc.). |

the target behaviors. Pressing a specific number once would mark the onset of the behavior, while pressing the same number again or the space bar on the keyboard would mark the offset of the behavior. The program is capable of recording up to 10 different behaviors and summarizing their respective frequencies and durations at the end of each session. This program can be used on any IBM compatible personal computer.

Observer Training and Interobserver-Reliability Checks

Three observers, the principal investigator, a supervisor from the Society of Manitobans with Disabilities, and a psychology undergraduate student were used to record real time data. Each observer was trained by the principal researcher. Training involved reviewing behavioral categories and observation systems, refining operational definitions of the target behaviors, and writing a closed-book exam on definitions of targeted behaviors and their respective codes. A score of 100% on the written exam was required before advancing to the second phase of training, during which the observers observed and recorded the target behaviors with the "Videocount" computer program.

The first four (pilot) sessions of the videotaped activities were used for training purposes and were scored under real time recording. Observers then discussed problems encountered during the observation, clarified grey areas and redefined target behaviors so as to minimize ambiguity. A minimum of 90% interobserver agreement with the principal researcher on 4 consecutive 30-min observations was required before observation and recording of experimental data commenced. Interobserver reliabilities on duration and frequency measures were calculated using duration and frequency ratios. These ratios were calculated by dividing the smaller total (duration or frequency in a session) by the larger total and then multiplying by 100 (Kazdin, 1982, p. 52).

To maximize interobserver reliability (IOR), IOR checks were conducted after every session throughout the project. The minimum interobserver agreement obtained was 90% while the maximum obtained was 100% (M = 94%).

Procedure

Participants' behaviors were videotaped during the free play sessions. A remote omni-directional

microphone was clipped onto the clothing of the focal participant to pick up verbal interactions between the participant and other individuals, e.g., teachers, peers. To minimize reactivity, the cameraperson pretended to videotape for two sessions to provide an adaptation period for the participants and other persons present in the observational setting. As soon as the videotaping of the first participant was completed, videotaping of the second participant began. Order of videotaping among the participants was determined randomly. A total of 12 30-min free play sessions was taped for each participant.

## Application of Suen and Ary Method to Chan (1992) Observations

Determining values of b and t. As no previous estimates of the values of the shortest behavioral bout (b) and shortest interresponse time (t) were available, the first four sessions were used as pilot sessions from which the values of b and t were obtained. These values are shown in Tables 3 and 4. As shown in these Tables, the value of b varied from 2 to 132 s while the value of t varied from .7 to 78 s with most of the values being less than 10 s. The long duration (131.98

Table 3

Duration of Shortest Behavioral Bout (b) and Shortest Interresponse Time (t) in seconds from Pilot Sessions of Ronnie.

| | Target Behaviors | | | | | |
|---|---|---|---|---|---|---|
| | Sharing and Cooperating | | Solitary Play | | Parallel Play | |
| Sessions | b | t | b | t | b | t |
| 1 | 1.978 | 12.857 | N/A | N/A | 2.143 | 1.319 |
| 2 | 3.022 | 2.527 | 3.681 | 3.681 | 4.670 | 2.308 |
| 3 | 9.670 | 2.198 | 4.615 | 3.297 | 6.044 | 2.473 |
| 4 | 131.98 | 61.044 | 35.989 | N/A | 4.121 | 12.527 |
| M | 36.642 | 19.657 | 11.071 | 3.489 | 4.245 | 4.657 |
| SD | 63.596 | 28.032 | 16.731 | 4.137 | 1.618 | 5.272 |
| .5M | 18.321 | 9.829 | 5.536 | 2.631 | 2.123 | 2.339 |

Note. N/A = not applicable, behavior did not occur

Table 4

Duration of Shortest Behavioral Bout (b) and Shortest Interresponse Time (t) in seconds from Pilot Sessions of Celine.

| | Target Behaviors | | | | | |
|---|---|---|---|---|---|---|
| | Sharing and Cooperating | | Solitary Play | | Parallel Play | |
| Sessions | b | t | b | t | b | t |
| 1 | 1.429 | 5.934 | 2.473 | 11.319 | 2.198 | 1.429 |
| 2 | 1.813 | 0.934 | 5.989 | 2.527 | N/A | N/A |
| 3 | 8.626 | 77.747 | 4.505 | 4.505 | 2.967 | 1.593 |
| 4 | 2.692 | 5.385 | 3.352 | 2.692 | 9.890 | 2.692 |
| M | 3.640 | 22.500 | 4.080 | 5.261 | 5.018 | 1.904 |
| SD | 3.366 | 36.900 | 1.521 | 4.137 | 4.236 | 0.687 |
| .5M | 1.820 | 11.250 | 2.040 | 2.631 | 2.509 | 0.952 |

Note. N/A = not applicable, behavior did not occur

s) of sharing and cooperating (Table 3) occurred when the children pretended they were camping. The additional eight free play sessions were used to provide the test data.

Optimally appropriate interval lengths for each behavior were selected according to the procedures suggested by Suen and Ary (1986a). As shown in Tables 3 and 4, the mean values of $\underline{b}$ and $\underline{t}$ are quite low. These mean values would have been much lower had they not been averaged with the longer $\underline{b}$ and $\underline{t}$ in the fourth pilot session. Since $\underline{b}$ and $\underline{t}$ values, or $.5\underline{b}$ and $.5\underline{t}$ must meet the requirement of being shorter than $\underline{i}$ to achieve a low probability of error, it would be necessary to have an interval of less than 3 s in the present case.

Unlike scoring of behaviors generated by electromechanical devices (e.g., pen deflection), for which occurrence (or non-occurrence) can be determined easily and instantly,

scoring of social behaviors requires taking into consideration the social settings and context of the participants. It is not likely that observers can determine the type of social behaviors the participants are engaging in simply by a brief glance. Hence, for those behaviors with very short durations of b and t (e.g., less than 6 s), Suen and Ary (1986b) suggested that the shortest interval that can realistically be used by an observer be selected.

From my experience in scoring the pilot sessions and from reviewing other research literature which used time sampling procedures to record similar play behaviors (e.g., Goldstein & Cisar, 1992; Rolfe & Crossley, 1991; Wintre, 1989), the shortest realistic interval that allows the observer to assess the social context and to score play behaviors is about 5 s. Hence the shortest realistic interval length was selected as 5-s for scoring the test sessions. Because momentary time sampling requires the observer to scan and score the behavior at specified moments and to ignore the behavior for the rest of the interval, an additional 5 s was added to make up the shortest realistic interval (scoring to be conducted during the

first 5 s of the 10-s interval).

In order to provide comparison intervals for those specified by Suen and Ary's method, longer sampling intervals which would be inappropriate according to Suen and Ary's procedures were also employed to score the target behaviors. Criteria for the selection of inappropriate intervals were that: (a) these longer intervals had been employed in other studies involving similarly defined play behavior, and (b) the intervals yielded a high probability of having measurement errors according to the calculations of Suen and Ary (1986a). A review of recent literatures (1990-present) indicated that a 30-s interval for partial-, and whole-interval (Bay-Hinitz, Peterson, & Quilitch, 1994; Kaiser, Snyder, & Rogers, 1995), and a 60-s interval for momentary time sampling (Hundert & Hopkins, 1992) met the first criterion. As shown in Tables 5 and 6, the high probability of error (1.0) associated with these intervals met the second criteria. Hence, the inappropriate intervals were 30 s for partial- (PI-30), and whole-interval (WI-30), and as 60 s for momentary time sampling (MTS-60). Tables 5 and 6 also show that intervals that met Suen and Ary's criteria (optimal

Table 5

Probabilities of Having Measurement Errors as a Function of Sampling Interval Appropriateness (Ronnie).

| | Target Behaviors | | | | | |
|---|---|---|---|---|---|---|
| | Sharing and Cooperating | | Solitary Play | | Parallel Play | |
| | i | p(e) | i | p(e) | i | p(e) |
| Appropriate Intervals | | | | | | |
| P.I | 5* | 0.0750 | 5** | 0.9384 | 5** | 0.9643 |
| W.I | 5* | 0.5830 | 5** | 0.9510 | 5** | 0.9796 |
| MTS | 10* | 0.0304 | 10** | 0.4599 | 10** | 0.9972 |
| Inappropriate Intervals | | | | | | |
| P.I | 30 | 1.0000 | 30 | 1.0000 | 30 | 1.0000 |
| W.I | 30 | 1.0000 | 30 | 1.0000 | 30 | 1.0000 |
| MTS | 60 | 1.0000 | 60 | 1.0000 | 60 | 1.0000 |

Note. i = interval length (in seconds) selected; p(e) = probability of having measurement errors; P.I = partial-interval recording procedure; W.I = whole-interval recording procedure; MTS = momentary time sampling procedure. * = optimal interval length. ** = shortest realistic interval length.

Table 6

Probabilities of Having Measurement Errors as a Function of Sampling Interval Appropriateness (Celine)

| | Target Behaviors | | | | | |
|---|---|---|---|---|---|---|
| | Sharing and Cooperating | | Solitary Play | | Parallel Play | |
| | i | p(e) | i | p(e) | i | p(e) |
| **Appropriate Intervals** | | | | | | |
| P.I | 5** | 0.6060 | 5** | 1.0000 | 5** | 1.0000 |
| W.I | 5** | 0.9834 | 5** | 1.0000 | 5** | 1.0000 |
| MTS | 10** | 0.9987 | 10** | 1.0000 | 10** | 1.0000 |
| **Inappropriate Intervals** | | | | | | |
| P.I | 30 | 1.0000 | 30 | 1.0000 | 30 | 1.0000 |
| W.I | 30 | 1.0000 | 30 | 1.0000 | 30 | 1.0000 |
| MTS | 60 | 1.0000 | 60 | 1.0000 | 60 | 1.0000 |

Note. i = interval length (in seconds) selected; p(e) = probability of having measurement errors; P.I = partial-interval recording procedure; W.I = whole-interval recording procedure; MTS = momentary time sampling procedure. * = optimal interval length. ** = shortest realistic interval length.

intervals) had low probability of errors. Shortest realistic intervals were considered appropriate although they sometimes yielded a substantial probability of having measurement errors.

Procedures for applying sampling methods. The videotaped experimental free play sessions were first scored under real time recording. Before the scoring of each session began, the principal researcher briefed the observers with information that might enhance the coding of targeted behaviors. For example, the principal researcher might inform the observers that the session began with the children playing under the theme of "shopping in a supermarket". Such information enabled the observer to code the behavior as "sharing and cooperating" as opposed to "solitary play".

Real time recording was conducted by pressing the numbers on the computer keyboard which represented the corresponding social behaviors (see Table 2). The "Videocount" program automatically recorded and stored the frequencies and durations of each behavior. At the completion of each session, a detailed record that included the onset and offset times and the durations of each behavioral episode in units of seconds,

together with a summary of total frequency and duration of each behavior scored was generated via the "Videocount" computer program. A sample record is provided in Appendix A.

With sampling procedures, each of the 30-min sessions were divided into smaller intervals of equal length (i). The obtained real time data (on computer printouts) were then manually scored under the rules of partial-interval, whole-interval, and momentary time sampling procedures (Suen & Ary, 1986b).

The sampling procedures were not used to score the original tapes by observers but were applied to the real time data record, as described above. This approach avoided the possibility of having observer error confounding the comparison of these sampling procedures. Any difference in the results produced by real time recording and the sampling procedures therefore can be attributed to the difference between the recording procedures.

Duration and Frequency Estimates.

The methods for duration and frequency estimation suggested by Suen and Ary (1986a) were employed.

Duration estimates were expressed as proportions. For real time recording, the proportion of time the target behavior occurred was obtained by dividing the cumulative time the behavior occurred during the session by the session length. For whole-interval, partial-interval, and momentary time sampling procedures, the proportion of time the target behavior occurred was estimated by the ratio between the number of intervals scored in accordance with the code (see Table 2) and the total number of intervals in a session. Frequency was estimated by counting the number of intervals scored in accordance with the code (see Table 2) that were immediately preceded by a "zero". If the first interval of the chain was scored "one", the frequency count would be adjusted by +1. The frequency estimate was expressed in absolute terms (i.e., not as a percentage).

## Results

Duration and frequency data collected under the various sampling procedures were first compared to the data obtained from real-time recording using a discrepancy index (Green et al., 1982). Algebraically, the discrepancy index can be expressed as [($D_s$ - $D_r$) /

$D_r$] x 100, where $D_s$ is the duration estimate obtained under the sampling procedures of interest and $D_r$ is the duration measure obtained under real time recording. The discrepancy index for frequency data was calculated similarly. A low score indicates small discrepancy (accuracy), while a high score indicates large discrepancy (inaccuracy). In addition, a positive score indicates over-estimation while a negative score indicates under-estimation. The discrepancy indices were calculated session by session and can be found in Appendices B (Ronnie) and C (Celine), along with the respective duration and frequency raw data.
For the following data presentation duration data always will be reported first, followed by frequency data.

The discrepancy indices for duration and frequency measures were entered into separate repeated measures analysis of variance (ANOVA). The design was a 2 x 3 x 3 repeated measures ANOVA with interval (appropriate vs inappropriate), sampling methods (PI, WI, MTS), and social behaviors (sharing, solitary play, parallel play) as the three factors. Summary tables of these ANOVAs can be found in Appendix D.

Of particular interest is the significant two factor interactions between interval and sampling method for each dependent measure and participant. For Ronnie $F(2.28) = 10.03$, $p < .0005$; for Celine $F(2,28) = 4.71$, $p < .017$. As there was no statistically reliable three-way interaction between interval, sampling methods, and behavior for either response measure or participant, the data were averaged across behavior. As shown in the top panel of Figure 1, the difference between appropriate and inappropriate intervals varies with sampling method. To illustrate, PI with an appropriate interval under-estimated duration by approximately 2%; while with an inappropriate interval, it over-estimated duration by approximately 12%. In contrast, WI with an appropriate interval over-estimated duration by approximately 3%; while with an inappropriate interval, it under-estimated duration by approximately 35%. MTS with an appropriate interval over-estimated duration by approximately 10% while; with an inappropriate interval, duration is over-estimated by approximately 27%. Celine's duration data provides a relatively exact replication, except

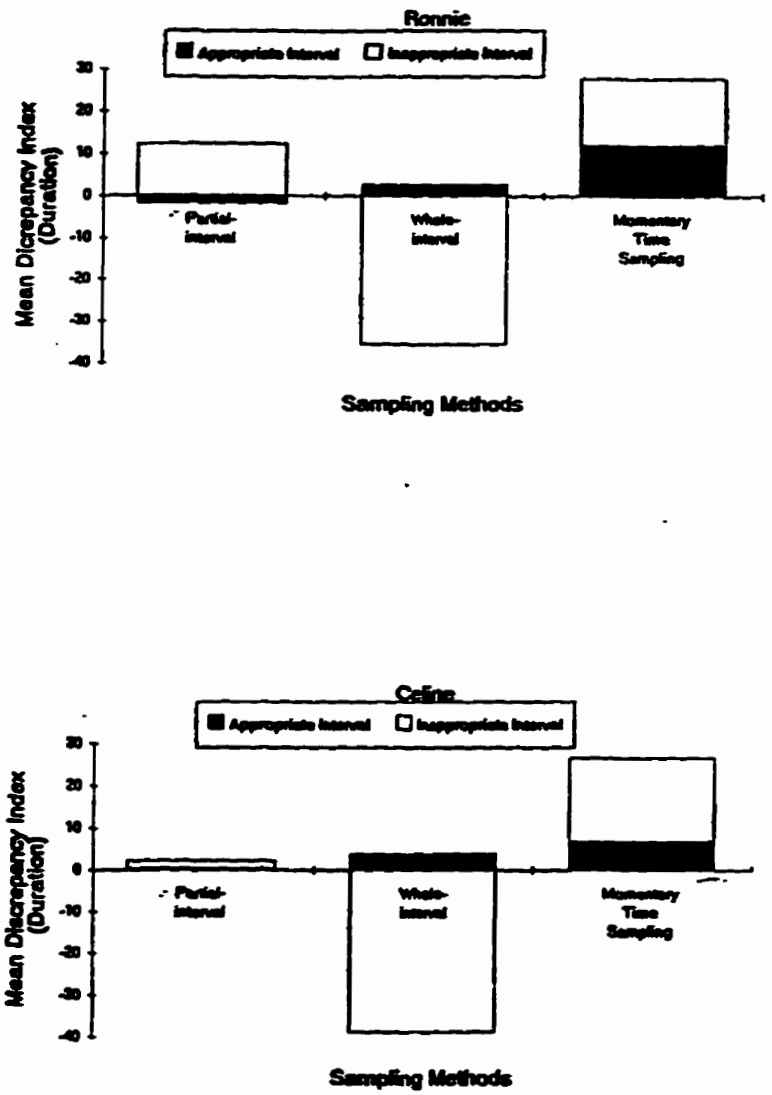



Figure 1. Duration estimates as a function of appropriateness of sampling intervals and sampling methods.

that there is very little error for PI.

The small error in PI for Ronnie and Celine is due to the cancellation of over- and under-estimations, so these data were reanalysed using absolute values of the descrepancy indices. As shown in Figure 2, the absolute values reveal a much larger difference between appropriate and inappropriate intervals. For PI, the mean absolute index is 26% with the appropriate interval and 168% with the inappropriate interval for Ronnie; and 33% with the appropriate interval and 241% with appropriate interval for Celine.

Turning to frequency measures, a statistically reliable interaction between interval and sampling method also emerged: for Ronnie $F(2,28) = 5.58$, $p < .0005$; for Celine $F(2,28) = 10.11$, $p < .001$. As shown in Figure 3, the difference between appropriate and inappropriate interval again varies with sampling method. To illustrate for PI, with an appropriate interval, frequency is under-estimated by 10%, while with an inappropriate interval, it is under-estimated by 40%. For WI with an appropriate interval, frequency is under-estimated by approximately

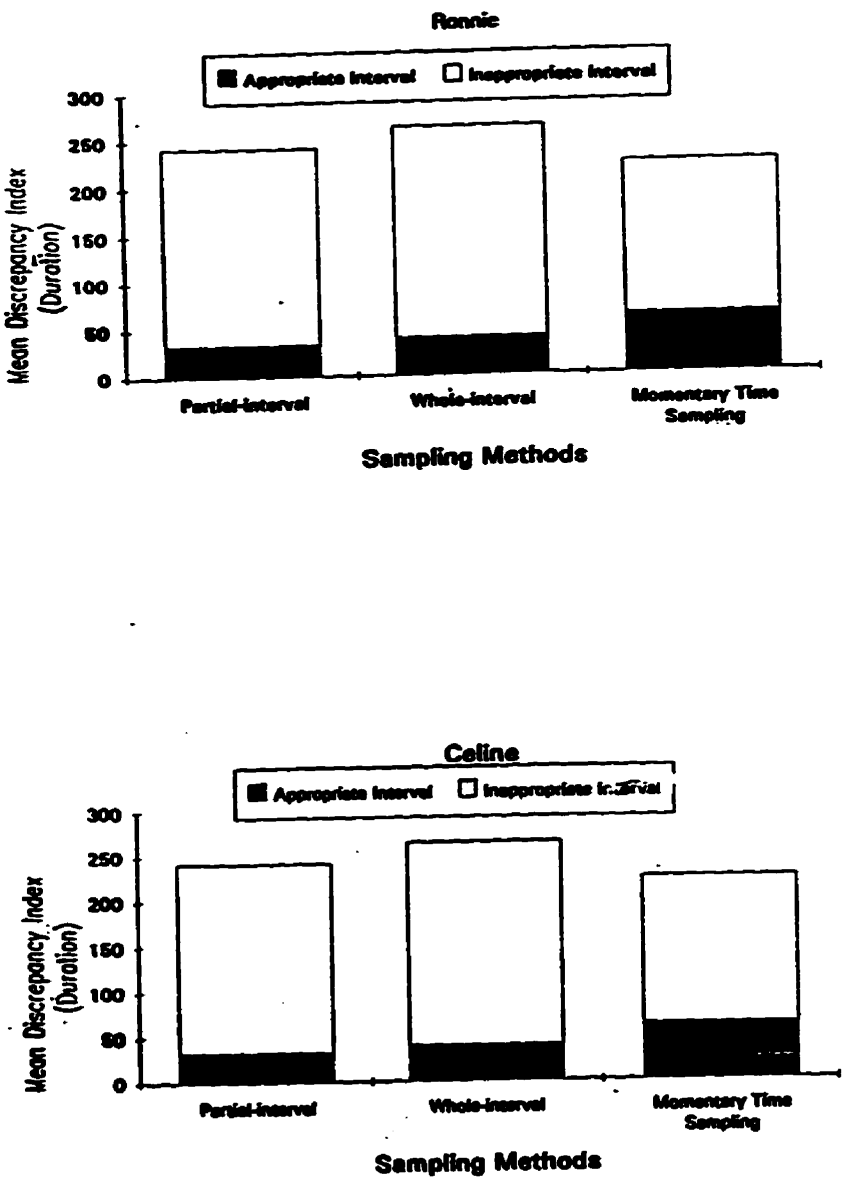



Figure 2. Duration estimates as a function of sampling intervals and sampling methods, using absolute discrepancy indices.

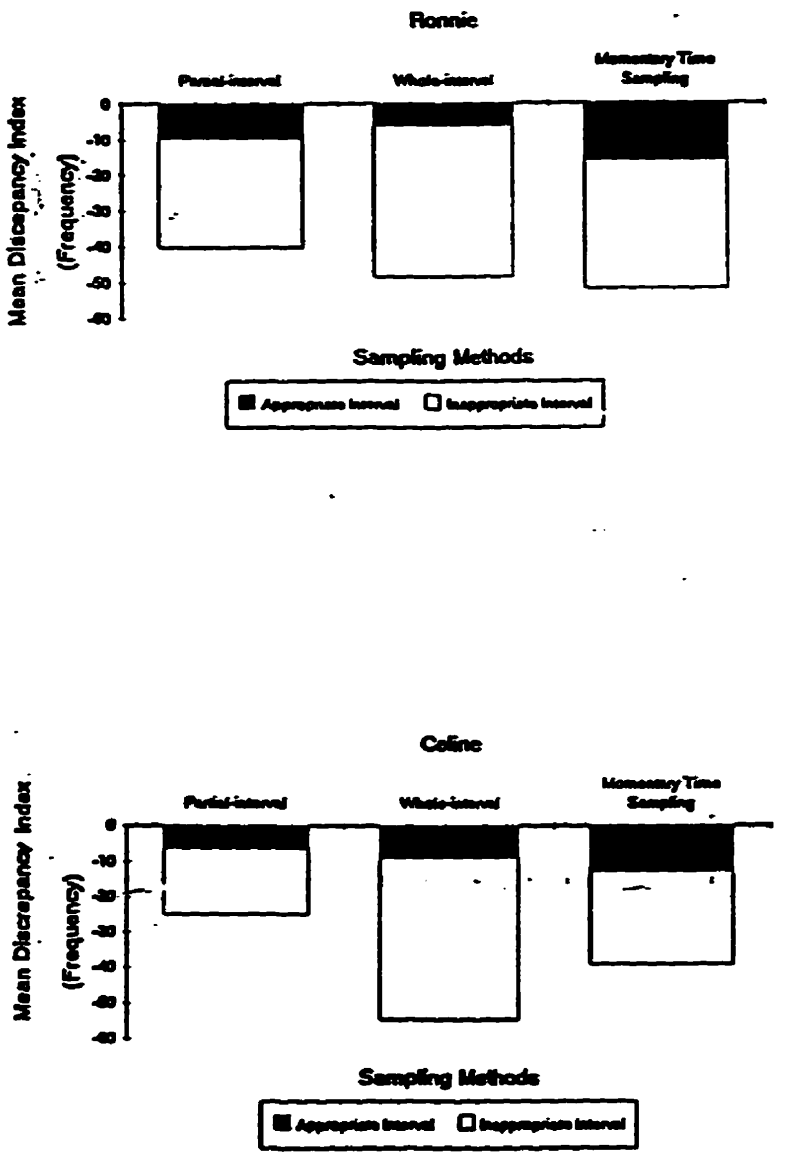



**Figure 3.** Frequency estimates as a function of appropriateness of sampling intervals and sampling methods.

7%; while with an inappropriate interval, frequency is under-estimated by approximately 48%. For MTS, with an appropriate interval, frequency is under-estimated by approximately 17%, while with an inappropriate interval, frequency is under-estimated by approximately 50%.

For Ronnie's data a statistically reliable interaction occurred between sampling method and type of social behavior for both dependent measures, and between sampling interval and type of social behavior for frequency measure only. As these were only of peripheral interest, the interested reader can find the relevant data in Appendix E. Of more central interest is the matter of whether estimates of duration or frequency, with either appropriate or inappropriate intervals, differed reliably from those of real-time recording. To examine this issue, discrepancy indices for each sampling method and interval were tested via paired t-tests with alpha = .05. The discrepancy index itself represents the difference (d) between real-time and the sampling data, and thus tested the null hypothesis of zero difference. As shown in Table 7, starting with Ronnie, duration estimates from PI are

Table 7

Comparisons of Estimates of Duration for Sampling and Real Time Measures

| Recording Procedures | Intervals | Participants | |
|---|---|---|---|
| | | Ronnie | Celine |
| | | t(7) | t(7) |
| | Appropriate | | |
| PI | 5 | 2.003 | -.120 |
| WI | 5 | -2.920* | -1.873 |
| MTS | 10 | -3.802** | -2.999* |
| | Inappropriate | | |
| PI | 30 | -1.846 | -1.025 |
| WI | 30 | 4.265** | 3.488* |
| MTS | 60 | -2.656* | -2.447* |

Note. PI = partial-interval recording; WI-5 = whole-interval recording; MTS = momentary time sampling. Intervals are in seconds.
*p < .05. **p < .01

not significantly different from real-time measures whether appropriate intervals are used or not. In contrast, duration estimates from WI and MTS are significantly different from real-time measures whether appropriate intervals are used or not. Celine's data replicate Ronnie's save for the single exception of no significant difference with WI when appropriate intervals are used. Turning to frequency estimates (Table 8), starting with Ronnie, significant differences were found in all sampling methods (PI, WI, MTS) with either intervals (appropriate and inappropriate). Again, Celine's data replicate Ronnie's save for PI where no statistically reliable difference was found with an appropriate interval.

## Discussion

The main objective of my research was to test the statistical procedures proposed by Suen and Ary (1986a) for determining optimal interval length in direct observation sampling procedures. The hypothesis was that if interval lengths were selected on the basis of the sampled, shortest behavioral bouts and the shortest interresponse times according to the method of Suen and Ary (1986a), the probability of having measurement

Table 8

Comparisons of Estimates of Frequencies for Sampling and Real Time Measures

| Recording Procedures | Intervals | Participants | |
|---|---|---|---|
| | | Ronnie | Celine |
| | | $t(7)$ | $t(7)$ |
| | Appropriate | | |
| PI | 5 | 6.794*** | 2.335 |
| WI | 5 | 3.064** | 6.911*** |
| MTS | 10 | 8.276**** | 5.202** |
| | Inappropriate | | |
| PI | 30 | 9.723**** | 5.524**** |
| WI | 30 | 10.023**** | 6.778**** |
| MTS | 60 | 8.327**** | 11.074**** |

Note. PI = partial-interval recording; WI-5 = whole-interval recording; MTS = momentary time sampling. Intervals are in seconds.
**$p$ < .01. ***$p$ < .001 ****$p$ < .0001

errors in partial-interval, whole-interval, and momentary time sampling would be minimized. Sampling intervals with low probability of having measurement errors (appropriate intervals) were expected to yield results that closely approximated real-time data. In contrast, sampling intervals with high probability of having measurement errors (inappropriate intervals) were expected to yield results that differed from real-time data.

The results of my research demonstrate that, consistent with Suen and Ary's model, the mean discrepancy indices were considerably lower with appropriate sampling intervals than with inappropriate intervals. Hence, the selection of an appropriate interval is of critical importance for accurate assessment. For example, if a researcher wishes to assess the effectiveness of a treatment program aimed at increasing children's social behavior in an integrated classroom, and the same sampling method is used to collect treatment data for all children, the researcher may find treatment success (increase in duration of play behavior) for one child and treatment failure for another child. The different findings may

be a reflection of the sampling interval being appropriate in the case of the first child (whose b and t satisfy the conditions required of accurate estimation) but not in the case of the second child, resulting in substantial error of measurement for the second child that masks the treatment effect.

Researchers such as Altmann (1974) and Kraemer (1979) have suggested that PI be abandoned because it reflects both duration and frequency. In contrast, Suen and Ary (1983) advocated PI to be the method of choice if researchers wish the sampling results to reflect a combination of both duration and frequency. The apparent inadequacy of PI perceived by Altmann and Kraemer may be a reflection of the Suen and Ary's required conditions not being met rather than a reflection of any inherent inadequacy. My research results suggest that if the interval length is close to the "ideal" interval derived from Suen and Ary's model, PI is capable of yielding duration data that are similar to real time measures.

Duration estimates yielded by MTS were found to be inaccurate regardless of whether appropriate intervals were used or not. This finding, however, should not be

taken as an indication of inadequacy in MTS because the predicted probability of error was very high with both intervals. The higher discrepancy index scores for MTS in comparison to PI may simply be a reflection of the intervals used being farther from the "ideal" interval derived from Suen and Ary' model when compared with the intervals used for PI sampling. My results are consistent with those of other researchers (e.g., Ary & Suen, 1983; Powell et al., 1977)' that found WI to systematically under-estimate duration.

The frequency estimates were, as predicted, biased. Although real time frequencies were consistently under-estimated by all sampling methods, those methods with inappropriate intervals under-estimated frequency by a much larger amount than those with appropriate intervals. This finding was consistent with other literature which found partial-interval, whole-interval, and momentary time sampling not to correspond well to the actual frequency of behavior (Altman & Wagner, 1970; Lentz, 1982; Saudargas & Lentz, 1986). As Suen and Ary (1984) remarked, when the conditions necessary for a correct frequency count are not met, frequency will likely be underestimated.

Based on the results of the current study, it appears that frequency estimates are more sensitive than duration estimates to violations of the conditions listed in Table 1. For example, frequency estimates from all sampling methods regardless of intervals were significantly different from real time measures save for the single exception of no significant difference with WI when appropriate intervals were used. Researchers interested in an accurate frequency count of a behavior should consider using real time recording only, or use the shortest realistic interval.

Due to the extremely short duration of most of the shortest behavioral bouts and shortest interresponse times in the pilot sessions, the selection of interval length via these statistical procedures was feasible for only one of the three target behaviors (sharing and cooperating) of Participant 1, and was not feasible for any of the target behaviors of Participant 2. Consequently, the shortest realistic interval length was employed. Because the probability of having measurement error was very high even with the shortest realistic interval, researchers may be discouraged from using the shortest realistic intervals. However, a

high probability of error, according to Suen and Ary's model, simply indicates that it is very likely that at least one behavioral bout or one interresponse time would be shorter than the selected interval. Their model does not predict the actual amount of error associated with the calculated probability of measurement error. The implication is that, after applying Suen and Ary's model, researchers may notice that it is very likely that error would occur when the behavior of interest is being scored under a certain sampling method, using a particular interval, but researchers may not have enough information to decide whether the interval should be used or not, because Suen and Ary's model does not offer information on how large these errors might be. For example, with Celine's pilot data on solitary and parallel play, the probability of error is 1.0 for all three sampling methods whether appropriate intervals are used or not.

From a practical point of view, if sampling methods were to be used, it would make sense to use the longer (inappropriate) intervals given that both intervals have the same probability of error. However, as shown by my results, using the longer

(inappropriate) interval results in a much larger discrepancy (e.g., 35%) with real-time data than using the shorter (appropriate) intervals (e.g., 4%). The lack of information about the amount of error associated with the appropriate interval length may prompt researchers to use inappropriate intervals without realizing that a significant amount of information may be lost. Hence, it is important for researchers to use appropriate, or shortest realistic intervals (as defined by Suen and Ary's model), even if the probability of error is as high as that of longer intervals.

The stringent requirements of Suen and Ary's model result in a high probability of error even with short intervals. The model predicted a high probability of error even with the shortest realistic interval of 5 s. In fact, their model predicted a probability of error of one even with intervals as short as 3 s (as in the case of the pilot data in this research). While the probability of measurement errors associated with certain interval lengths can be assessed by applying Suen and Ary's procedures, their procedures do not offer information on how large those errors will be.

An enhancement to Suen and Ary's model would be to predict the amount of error associated with a sampling interval. If the amount of error is acceptable to the researcher, the interval may still be applicable even if the probability of error is very high. Based on the results of my research, I would assume that measurement errors from sampling methods using appropriate intervals would be much less than those from sampling methods using inappropriate intervals. The theoretical assumption is that the farther the selected interval is from the "ideal" interval derived from Suen and Ary's model, the larger the discrepancy index will be. Development of procedures capable of predicting the size of measurement errors associated with various sampling intervals would greatly enhance the usefulness of Suen and Ary's model for researchers. In the meantime, researchers need to be careful to select an appropriate interval, using Suen and Ary's model. Afterall, accurate measurement is the keystone of valid assessment and treatment.

## References

Altmann, J. (1974). Observational study of behavior sampling methods. Behavior, 49, 31-45.

Altmann, A. S., & Wagner, S. S. (1970). Estimating rates of behavior from Hanson frequencies. Primates, 11, 181-183.

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

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.

Bakeman, R., & Gottman, J. M. (1987). Applying observational methods: A systematic review. In J. D. Osofsky (Ed.), Handbook of infant development (2nd ed., pp. 818-854). New York: Wiley.

Bass, R. F. (1987). Computer-assisting observer training. Journal of Applied Behavior Analysis, 20, 83-88.

Bay-Hinitz, A. K., Peterson, R. F., & Quilitch, H. R. (1994). Cooperative games: a way to modify

aggressive and cooperative behaviors in young children. Journal of Applied Behavior Analysis, 27, 435-446.

Bijou, S. W., Peterson, R. F., & Ault, M. H. (1968). A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. Journal of Applied Behavior Analysis, 1, 175-191.

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.

Celhoffer, L., Boukydis, C., Minde, K., & Muir, E. (1977). The DAR-11 event recorder: a portable high-speed digital cassette system with direct computer access. Behavior Research Methods and Instrumentation, 9, 442-446.

Chan, D. (1992). Partial-, whole-interval and continuous recording of social behaviors in preschool children. Honours Thesis, University of Manitoba, Winnipeg.

Cochran, W. G. (1977). Sampling Techniques (r ed.). New York: John Wiley & Sons.

Cullen, J. (1993). Preschool children's use and perception of outdoor play areas. Early Child Development and Care, 89, 45-56.

Edwards, R., Kearns, K., & Tingstrom, D. H. (1991). Accuracy of long momentary time-sampling intervals: Effects of errors in the timing of observations. Journal of Psychoeducational Assessment, 9, 160-165.

Fitzpatrick, L. J. (1977). Automated data collection for observed events. Behavior Research Methods and Instrumentation, 9, 447-451.

Foster, S. L., Bell-Dolen, D. L., Burge, D. A. (1988). Behavioral observation. In A. S. Bellack & M. Hersen (Ed.). Behavioral Assessment (r ed.).

Goldstein, H., & Cisar, C. L. (1992). Promoting interaction during sociodramatic play:Teaching scripts to typical preschoolers and classmates with disabilities. Journal of Applied Behavior Analysis, 25, 265-280.

Green, 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 technique. Journal of Behavioral Assessment, 4, 103-110.

Guralnick, M. J. (1990). Social competence and early intervention. Journal of Early Intervention, 14, 3-14.

Haring, T. G., & Breen, C. G. (1989). Units of analysis of social interaction outcomes in supported education. Journal of the Association for Persons with Severe Handicaps, 14, 255-262.

Haring, T. G. , & Breen, C. G. (1992). A peer-mediated social network intervention to enhance the social integration of persons with moderate and severe disabilities. Journal of Applied Behavior Analysis, 25, 319-333.

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

Hay, W. L. (1973). Statistics for the social sciences.. New York: Holt, Rinehart, and Winston.

Hundert, J., & Hopkins, B. (1992). Training supervisors in a collaborative team approach to promote peer interaction of children with disabilities in integrated preschool. Journal of Applied Behavior Analysis, 25, 385-400.

Hutt, S. J., & Hutt, C. (1970). Direct Observation and Measurement of Behavior. Springfield, Ill: Thomas.

Kaiser, J., Snyder, T. A., & Rogers, C. S. (1995). Adult choice of toys affects children's prosocial and antisocial behavior. Early Child Development and Care, 111, 181-193.

Kazdin, A. E. (1982). Single-case research design. Methods for clinical and applied settings. New York: Oxford University Press.

Kelly, M. B. (1977). A review of the observational data-collection and reliability procedures reported in the Journal of Applied Behavior Analysis. Journal of Applied Behavior Analysis, 10, 97-101.

Klesges, R. C., Woolfrey, J., & Vollmer, J. (1985). An evaluation of the reliability of time sampling versus continuous observation data collection.

Journal of Behavioral Therapy & Experimental Psychiatry, 16, 303-307.

Kraemer, H. (1979). One-zero sampling in the study of primate behavior. Primates, 20, 237-244.

Lentz, F. E. (1982). An empirical examination of the utility of partial interval and momentary sampling as measurements of behavior. Unpublished Doctoral Dissertation, University of Tennessee.

Lipinski, D., & Nelson, R. (1974). Problems in the use of naturalistic observation as a means of behavioral assessment. Behavior Therapy, 5, 341-351.

Lyytinen, P. (1991). Peer interaction in children's dyadic play. Early Child Development and Care, 71, 105-115.

Mansell, J. (1985). Time sampling and measurement error: The effect of interval length and sampling pattern. Journal of Behavioral Therapy and Experimental Psychiatry, 16, 245-251.

Parten, M. B. (1932). Social participation among pre-school children. Journal of Abnormal and Social Psychology, 27, 243-269.

Powell, J. (1984a). On the misrepresentation of behavioral realities by a widely practiced direct

observation procedure: Partial interval (one-zero) sampling. Behavioral Assessment, 6, 209-219.

Powell, J. (1984b). Some empirical justification for a modest proposal regarding data acquisition via intermittent direct observation. Journal of Behavioral Assessment, 6, 71-80.

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

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

Quera, V. (1991). A generalized technique to estimate frequency and duration in time sampling. Behavioral Assessment, 12, 409-424.

Repp, A. C. (1983). Teaching the mentally retarded. Englewood Cliffs, NJ: Prentice Hall.

Repp, A. C., Barton, L., & Brulle , A. (1982). Naturalistic observations of retarded persons. Paper given at Lake Wilderness Conference on the impact of residential environments on retarded persons and their careproviders.

Repp, A. C., Nieminen, G. S., Olinger, E., & Brusca, R. (1988). Direction observation: Factors affecting the accuracy of observers. Exceptional Children, 55, 29-36.

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.

Rhine, R. J., & Linville, A. K. (1980). Properties of one-zero scores in observational studies of primate social behavior: The effect of assumptions on empirical analysis. Primates, 21, 111-122.

Rojahn, J., & Kanoy, R. (1985). Toward an empirically based parameter selection for time-sampling observation systems. Journal of Psychopathology and Behavioral Assessment, 7, 99-120.

Rolfe, S. A., & Crossley, S. A. (1991). The effects of prior experience in early childhood programs

and play setting on social behavior and play in preschoolers. Early Childhood and Care, 72, 23-37.

Sanson-Fisher, R. W., Poole A. D., Small, G. A., & Fleming, I. (1979). Data acquisition in real time: an improved system for naturalistic observation. Behavior Therapy, 10, 543-554.

Sanson-Fisher, R. W., Poole, A. D., & Dunn, J. (1980). An empirical method for determining an appropriate interval length for recording behavior. Journal of Applied Behavior Analysis, 13, 493-500.

Saudargas, R. A., & Lentz, F. E. (1986). Estimating percent of time and rate via direct observation: A suggested observational procedure and format. School Psychology Review, 15, 36-48.

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.

Slate, J. R., & Saudargas, R. A. (1986). Differences in learning disabled and average students' classroom behaviors. Learning Disability Quarterly, 9, 61-67.

Springer, B., Brown, T., & Duncan, P. (1981). Current measurement in applied behavior analysis. The Behavior Analyst, 4, 19-31.

Struffy, J., & Struffy, C. (1985). Friendship and our children. Journal of the Association for Persons with Severe Handicaps, 10, 224-227.

Suen, H. K., & Ary, D. (1984). Variables influencing one-zero and instantaneous time sampling outcomes. Primates, 25, 89-94.

Suen, H. K., & Ary, D. (1986a). Poisson cumulative probabilities of systematic errors in single-participant and multiple-participant time sampling. Behavioral Assessment, 8, 155-169.

Suen, H. K., & Ary, D. (1986b). A post hoc procedure for systematic errors in time-sampling duration estimates. Journal of Psychopathology and Behavioral Assessment, 8, 31-38.

Suen, H. K., & Ary, D. (1989). Analyzing quantitative behavioral observation data. Hilldale, NJ: LEA.

Wright, H. F. (1960). Observational child study. In P. H. Mussen (Ed.), Handbook of Research Methods in Child Development. New York, NY: Wiley

Wasik, B. H. (1984). Clinical application of direct behavioral observation: A look at the past and the future. In B. B. Lahey & A. E. Kazdin (Eds.), Advances in clinical child psychology (pp. 156-193). New York: Plenum Press.

Wasik, B. H. (1989). The systematic observation of children: Rediscovery and Advances. Behavioral Assessment, 11, 201-217.

Wintre, M. G. (1989). Changes in social play behavior as a function of preschool programs. Journal of Educational Research, 82, 294-301.

# Appendix A

## Sample of Real Time Data Printout from "Videocount"

| DATE/TIME | KEY | DURATION |
|---|---|---|
| Fri Jan 24 20:04:40 1992 | 9 | 39.451 |
| Fri Jan 24 20:05:10 1992 | 6 | 3.297 |
| Fri Jan 24 20:05:13 1992 | 2 | 30.934 |
| Fri Jan 24 20:05:44 1992 | 9 | 62.692 |
| Fri Jan 24 20:06:47 1992 | 2 | 13.077 |
| Fri Jan 24 20:07:00 1992 | 9 | 5.440 |
| Fri Jan 24 20:07:05 1992 | 4 | 2.198 |
| Fri Jan 24 20:07:07 1992 | 9 | 26.538 |
| Fri Jan 24 20:07:34 1992 | 4 | 6.813 |
| Fri Jan 24 20:07:41 1992 | 5 | 6.209 |
| Fri Jan 24 20:07:47 1992 | 4 | 21.648 |
| Fri Jan 24 20:08:09 1992 | 9 | 2.747 |
| Fri Jan 24 20:08:11 1992 | 8 | 3.626 |
| Fri Jan 24 20:08:15 1992 | 4 | 5.824 |
| Fri Jan 24 20:08:21 1992 | 9 | 20.934 |
| Fri Jan 24 20:08:42 1992 | 4 | 4.286 |
| Fri Jan 24 20:08:46 1992 | 9 | 12.967 |
| Fri Jan 24 20:08:59 1992 | 4 | 21.099 |
| Fri Jan 24 20:09:20 1992 | 9 | 66.978 |
| Fri Jan 24 20:10:27 1992 | 6 | 2.857 |
| Fri Jan 24 20:10:30 1992 | 9 | 26.703 |
| Fri Jan 24 20:10:57 1992 | 8 | 830.000 |
| Fri Jan 24 20:24:46 1992 | 5 | 310.659 |
| Fri Jan 24 20:29:57 1992 | 8 | 16.099 |
| Fri Jan 24 20:30:13 1992 | 5 | 39.725 |
| Fri Jan 24 20:30:53 1992 | 8 | 10.385 |
| Fri Jan 24 20:31:03 1992 | 5 | 107.527 |
| Fri Jan 24 20:32:51 1992 | 8 | 2.857 |
| Fri Jan 24 20:32:53 1992 | 5 | 91.044 |
| 20:34:24 | | |

| KEY | FRQ | DURATION | DURATNSQ | MEAN | VARIANCE | STANDDEV |
|---|---|---|---|---|---|---|
| 0 | 0 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 1 | 0 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 2 | 2 | 44.011 | 1127.922 | 22.005 | 159.439 | 12.627 |
| 3 | 0 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 4 | 6 | 61.868 | 1017.353 | 10.311 | 75.882 | 8.711 |
| 5 | 5 | 555.165 | 117977.023 | 111.033 | 14083.855 | 118.675 |
| 6 | 2 | 6.154 | 19.032 | 3.077 | 0.097 | 0.311 |
| 7 | 0 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 8 | 5 | 862.967 | 689288.312 | 172.593 | 135086.469 | 367.541 |
| 9 | 9 | 254.451 | 11344.587 | 28.272 | 518.836 | 22.778 |

Appendix B-1

Behavior Duration and Discrepancy Index of Sampling Estimates when Compared with Real Time Data (Ronnie)

**Sharing and Cooperating (code 2) Duration Data**

| Session | Real time | PI-5 | WI-5 | MTS- 10 | PI-30 | WI-30 | MTS- 60 |
|---|---|---|---|---|---|---|---|
| 1 | 0.1582 | 0.1528 | 0.1556 | 0.1667 | 0.1500 | 0.1500 | 0.2069 |
| 2 | 0.0289 | 0.0278 | 0.0306 | 0.0330 | 0.0164 | 0.0000 | 0.0323 |
| 3 | 0.0245 | 0.0222 | 0.0250 | 0.0278 | 0.0167 | 0.0000 | 0.0667 |
| 4 | 0.0218 | 0.0241 | 0.0267 | 0.0374 | 0.0317 | 0.0000 | 0.0313 |
| 5 | 0.0268 | 0.0219 | 0.0247 | 0.0328 | 0.0164 | 0.0000 | 0.0326 |
| 6 | 0.0932 | 0.0944 | 0.0944 | 0.1000 | 0.0820 | 0.0820 | 0.0968 |
| 7 | 0.4090 | 0.4083 | 0.4167 | 0.4167 | 0.4333 | 0.4000 | 0.4333 |
| 8 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| *Mean* | 0.0953 | 0.0939 | 0.0967 | 0.1018 | 0.0933 | 0.0790 | 0.1125 |
| *SD* | 0.1367 | 0.1366 | 0.1387 | 0.1378 | 0.1459 | 0.1411 | 0.1445 |

**Discrepancy Index**

| Session | | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS- 60 |
|---|---|---|---|---|---|---|---|
| 1 | | -3.41 | -1.64 | 5.37 | -5.18 | -5.18 | 30.78 |
| 2 | | -3.81 | 5.88 | 14.19 | -43.25 | -100.00 | 11.76 |
| 3 | | -9.39 | 2.04 | 13.47 | -31.84 | -100.00 | 172.24 |
| 4 | | 10.55 | 22.48 | 71.56 | 45.41 | -100.00 | 43.58 |
| 5 | | -18.28 | -7.84 | 22.39 | -38.81 | -100.00 | 21.64 |
| 6 | | 1.29 | 1.29 | 7.30 | -12.02 | -12.02 | 3.86 |
| 7 | | -0.17 | 1.88 | 1.88 | 5.94 | -2.20 | 5.94 |
| 8 | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | *Mean* | -2.90 | 3.01 | 17.02 | -9.97 | -52.43 | 36.23 |
| | *SD* | 8.40 | 8.79 | 23.22 | 28.94 | 50.97 | 56.91 |

Note. PI = partial-interval sampling; WI = whole-interval sampling; MTS = momentary time sampling. Number after the hyphen indicates sampling interval in seconds.

## Appendix B-2

### Behavior Duration and Discrepancy Index of Sampling Estimates When Compared with Real Time Data (Ronnie)

**Solitary Play (code4) Duration Data**

| Session | Real time | PI-5 | WI-5 | MTS- 10 | PI-30 | WI-30 | MTS- 60 |
|---|---|---|---|---|---|---|---|
| 1 | 0.3086 | 0.3111 | 0.3306 | 0.3611 | 0.3833 | 0.0833 | 0.3103 |
| 2 | 0.2348 | 0.2333 | 0.2389 | 0.2473 | 0.2459 | 0.1803 | 0.2581 |
| 3 | 0.0344 | 0.0361 | 0.0361 | 0.0444 | 0.0833 | 0.0000 | 0.0667 |
| 4 | 0.1103 | 0.1046 | 0.1126 | 0.1283 | 0.1429 | 0.0635 | 0.2188 |
| 5 | 0.3137 | 0.3151 | 0.3288 | 0.3497 | 0.3934 | 0.2131 | 0.3871 |
| 6 | 0.4660 | 0.4611 | 0.4694 | 0.4889 | 0.5574 | 0.4262 | 0.5161 |
| 7 | 0.2898 | 0.2861 | 0.2972 | 0.2944 | 0.3000 | 0.2667 | 0.3000 |
| 8 | 0.3231 | 0.3154 | 0.3342 | 0.3387 | 0.3548 | 0.2098 | 0.3548 |
| *Mean* | 0.2601 | 0.2579 | 0.2685 | 0.2816 | 0.3076 | 0.1804 | 0.3015 |
| *SD* | 0.1344 | 0.1335 | 0.1374 | 0.1406 | 0.1507 | 0.1339 | 0.1310 |

| Discrepancy Index Session | | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|---|
| 1 | | 0.81 | 7.13 | 17.01 | 24.21 | -73.01 | 0.55 |
| 2 | | -0.64 | 1.75 | 5.32 | 4.73 | -23.21 | 9.92 |
| 3 | | 4.94 | 4.94 | 29.07 | 142.15 | -100.00 | 93.90 |
| 4 | | -5.17 | 2.09 | 16.32 | 29.56 | -42.43 | 98.37 |
| 5 | | 0.45 | 4.81 | 11.48 | 25.41 | -32.07 | 23.40 |
| 6 | | -1.05 | 0.73 | 4.91 | 19.61 | -8.54 | 10.75 |
| 7 | | -1.26 | 2.55 | 1.59 | 3.52 | -7.97 | 3.52 |
| 8 | | -2.38 | 3.44 | 4.83 | 9.81 | -35.07 | 9.81 |
| | *Mean* | -0.54 | 3.43 | 11.32 | 32.37 | -40.29 | 31.28 |
| | *SD* | 2.89 | 2.09 | 9.15 | 45.42 | 31.80 | 40.60 |

Note. PI = partial-interval sampling; WI = whole-interval sampling; MTS = momentary time sampling. Number after hyphen indicates sampling interval in seconds.

## Appendix B-3

### Behavior Duration and Discrepancy Index of Sampling Estimates When Compared with Real Time Data (Ronnie)

**Parallel Play (code5) Duration Data**

| Session | Real time | PI-5 | WI-5 | MTS- 10 | PI-30 | WI-30 | MTS -60 |
|---|---|---|---|---|---|---|---|
| 1 | 0.2087 | 0.2056 | 0.2111 | 0.2167 | 0.2333 | 0.2333 | 0.2759 |
| 2 | 0.4092 | 0.4083 | 0.4194 | 0.4341 | 0.4263 | 0.4263 | 0.4839 |
| 3 | 0.3110 | 0.3109 | 0.3137 | 0.3222 | 0.3167 | 0.2667 | 0.3333 |
| 4 | 0.5695 | 0.5699 | 0.5726 | 0.5882 | 0.6190 | 0.5397 | 0.5625 |
| 5 | 0.4039 | 0.3934 | 0.4016 | 0.4153 | 0.4590 | 0.4098 | 0.4194 |
| 6 | 0.0767 | 0.0750 | 0.0805 | 0.0944 | 0.0820 | 0.0656 | 0.1290 |
| 7 | 0.1391 | 0.1278 | 0.1444 | 0.1500 | 0.1667 | 0.1000 | 0.1333 |
| 8 | 0.2612 | 0.2615 | 0.2642 | 0.2688 | 0.2742 | 0.2581 | 0.2581 |
| *Mean* | 0.2974 | 0.2941 | 0.3009 | 0.3112 | 0.3222 | 0.2874 | 0.3244 |
| *SD* | 0.1608 | 0.1620 | 0.1606 | 0.1632 | 0.1729 | 0.1632 | 0.1571 |

**Discrepancy Index**

| Session | | PI-5 | WI-5 | MTS-10 | PI30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|---|
| 1 | | -1.49 | 1.15 | 3.83 | 10.54 | 11.79 | 32.20 |
| 2 | | -0.22 | 2.49 | 6.09 | 4.01 | 4.18 | 18.26 |
| 3 | | -0.03 | 0.87 | 3.60 | 1.80 | -14.24 | 7.17 |
| 4 | | 0.07 | 0.54 | 3.28 | 8.00 | -5.23 | -1.23 |
| 5 | | -2.60 | -0.57 | 2.82 | 12.00 | 1.46 | 3.84 |
| 6 | | -2.22 | 4.95 | 23.06 | 6.46 | -14.47 | 68.19 |
| 7 | | -8.12 | 3.81 | 7.84 | 16.56 | -28.11 | -4.17 |
| 8 | | 0.11 | 1.15 | 2.91 | 4.74 | -1.19 | -1.19 |
| | *Mean* | -1.81 | 1.80 | 6.68 | 8.01 | -5.73 | 15.38 |
| | *SD* | 2.77 | 1.83 | 6.85 | 4.83 | 12.69 | 24.53 |

Note. PI = partial-interval sampling; WI = whole-interval sampling; MTS = momentary time sampling. Number after hyphen indicates sampling interval in seconds.

## Appendix B-4

### Behavior Frequency and Discrepancy Index of Sampling Estimates when Compared with Real Time Data (Ronnie)

**Sharing and Cooperating (code2) Frequency Data**

| Session | Real time | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|---|
| 1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 2 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 3 | 2 | 2 | 2 | 2 | 2 | 0 | 1 |
| 4 | 3 | 3 | 3 | 2 | 2 | 0 | 1 |
| 5 | 3 | 3 | 2 | 3 | 3 | 0 | 1 |
| 6 | 2 | 2 | 2 | 2 | 2 | 1 | 1 |
| 7 | 6 | 5 | 5 | 5 | 3 | 4 | 3 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Mean* | 2.625 | 2.500 | 2.375 | 2.375 | 2.125 | 1.125 | 1.500 |
| *SD* | 1.849 | 1.604 | 1.598 | 1.598 | 1.246 | 1.808 | 1.309 |

**Discrepancy Index**

| Session | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|
| 1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 2 | 0.00 | 0.00 | 0.00 | 0.00 | -100.00 | 0.00 |
| 3 | 0.00 | 0.00 | 0.00 | 0.00 | -100.00 | -50.00 |
| 4 | 0.00 | 0.00 | -33.33 | -33.33 | -100.00 | -66.67 |
| 5 | 0.00 | -33.33 | 0.00 | 0.00 | -100.00 | -66.67 |
| 6 | 0.00 | 0.00 | 0.00 | 0.00 | -50.00 | -50.00 |
| 7 | -16.67 | -16.67 | -16.67 | -50.00 | -33.33 | -50.00 |
| 8 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Mean* | -2.08 | -6.25 | -6.25 | -10.42 | -60.42 | -35.42 |
| *S.D* | 5.89 | 12.40 | 12.40 | 19.80 | 45.37 | 30.13 |

Note. PI = partial-interval sampling; WI = whole-interval sampling; MTS = momentary time sampling. Number after hyphen indicates sampling interval in seconds.

Appendix B-5

Behavior Frequency and Discrepancy Index of Sampling Estimates when Compared with Real Time Data (Ronnie)

Solitary Play (code 4) Frequency Data

| Session | Real time | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|---|
| 1 | 20 | 15 | 20 | 13 | 8 | 2 | 6 |
| 2 | 6 | 6 | 6 | 6 | 5 | 2 | 3 |
| 3 | 6 | 5 | 4 | 4 | 1 | 0 | 1 |
| 4 | 8 | 8 | 7 | 8 | 4 | 2 | 4 |
| 5 | 11 | 10 | 11 | 8 | 2 | 3 | 2 |
| 6 | 15 | 13 | 14 | 13 | 6 | 8 | 3 |
| 7 | 11 | 11 | 11 | 11 | 6 | 6 | 5 |
| 8 | 11 | 9 | 11 | 9 | 4 | 4 | 5 |
| *Mean* | 11.000 | 9.625 | 10.500 | 9.000 | 4.500 | 3.375 | 3.625 |
| *SD* | 4.721 | 3.378 | 5.043 | 10.286 | 2.268 | 2.560 | 1.685 |

Discrepancy Index

| Session | | PI-5 | WI-5 | MTS-10s | PI-30 | WI-30 | MTS-60s |
|---|---|---|---|---|---|---|---|
| 1 | | -25.00 | 0.00 | -35.00 | -60.00 | -90.00 | -70.00 |
| 2 | | 0.00 | 0.00 | 0.00 | -16.67 | -66.67 | -50.00 |
| 3 | | -16.67 | -33.33 | -33.33 | -83.33 | -100.00 | -83.33 |
| 4 | | 0.00 | -12.50 | 0.00 | -50.00 | -75.00 | -50.00 |
| 5 | | -9.09 | 0.00 | -27.27 | -81.82 | -72.73 | -81.82 |
| 6 | | -13.33 | -6.67 | -13.33 | -60.00 | -46.67 | -80.00 |
| 7 | | 0.00 | 0.00 | 0.00 | -45.45 | -45.45 | -54.55 |
| 8 | | -18.18 | 0.00 | -18.18 | -63.64 | -63.64 | -54.55 |
| | *Mean* | -10.28 | -6.56 | -15.89 | -57.61 | -70.02 | -65.53 |
| | *S.D* | 9.62 | 11.76 | 14.97 | 21.31 | 19.01 | 14.81 |

Note. PI = partial-interval sampling; WI = whole-interval sampling; MTS = momentary time sampling. Number after the hyphen indicates sampling interval in seconds.

## Appendix B-6

### Behavior Frequency and Discrepancy Index of Sampling Estimates when Compared with Real Time Data (Ronnie)

**Parallel Play (code 5) Frequency Data**

| Session | Real time | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|---|
| 1 | 6 | 6 | 6 | 6 | 5 | 4 | 5 |
| 2 | 12 | 11 | 12 | 9 | 5 | 8 | 4 |
| 3 | 5 | 4 | 5 | 3 | 2 | 3 | 3 |
| 4 | 12 | 9 | 10 | 9 | 4 | 8 | 4 |
| 5 | 11 | 9 | 10 | 8 | 3 | 6 | 3 |
| 6 | 6 | 5 | 5 | 5 | 3 | 2 | 1 |
| 7 | 8 | 7 | 8 | 5 | 3 | 2 | 3 |
| 8 | 3 | 2 | 3 | 2 | 2 | 2 | 2 |
| *Mean* | 7.875 | 6.625 | 7.375 | 5.875 | 4.500 | 3.375 | 3.125 |
| *SD* | 3.441 | 2.973 | 3.114 | 2.642 | 2.268 | 2.560 | 1.246 |

**Discrepancy Index**

| Session | | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|---|
| 1 | | 0.00 | 0.00 | 0.00 | -16.67 | -33.33 | -16.67 |
| 2 | | -8.33 | 0.00 | -25.00 | -58.33 | -33.33 | -66.67 |
| 3 | | -20.00 | 0.00 | -40.00 | -60.00 | -40.00 | -40.00 |
| 4 | | -25.00 | -16.67 | -25.00 | -66.67 | -33.33 | -66.67 |
| 5 | | -18.18 | -9.09 | -27.27 | -72.73 | -45.45 | -72.73 |
| 6 | | -16.67 | -16.67 | -16.67 | -50.00 | -66.67 | -83.33 |
| 7 | | -12.50 | 0.00 | -37.50 | -62.50 | -75.00 | -62.50 |
| 8 | | -33.33 | 0.00 | -33.33 | -33.33 | -33.33 | -33.33 |
| | *Mean* | -16.75 | -5.30 | -25.60 | -52.53 | -45.06 | -55.24 |
| | *S.D* | 10.18 | 7.68 | 12.80 | 18.74 | 16.64 | 22.71 |

Note. PI = partial-interval sampling; WI = whole-interval sampling; MTS = momentary time sampling. Number after hyphen indicates sampling interval in seconds.

Appendix C-1

Behavior Duration and Discrepancy Index of Sampling Estimates when Compared with Real Time Data (Celine)

**Sharing and Cooperating (code2) Duration Data**

| Session | Real time | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|---|
| 1 | 0.5766 | 0.5801 | 0.5897 | 0.5897 | 0.5962 | 0.5385 | 0.5385 |
| 2 | 0.0219 | 0.0219 | 0.0224 | 0.0256 | 0.0189 | 0.0000 | 0.0370 |
| 3 | 0.5625 | 0.5595 | 0.5757 | 0.5622 | 0.6129 | 0.5645 | 0.5806 |
| 4 | 0.4662 | 0.4587 | 0.4667 | 0.4840 | 0.4921 | 0.4603 | 0.5000 |
| 5 | 0.0112 | 0.0108 | 0.0134 | 0.0107 | 0.0159 | 0.0000 | 0.0313 |
| 6 | 0.2014 | 0.2020 | 0.2020 | 0.2092 | 0.2121 | 0.1970 | 0.2424 |
| 7 | 0.3239 | 0.3151 | 0.3385 | 0.3420 | 0.3750 | 0.1556 | 0.4688 |
| 8 | 0.0885 | 0.0854 | 0.0854 | 0.0915 | 0.0556 | 0.0370 | 0.1071 |
| *Mean* | 0.2815 | 0.2792 | 0.2867 | 0.2894 | 0.2973 | 0.2441 | 0.3132 |
| *SD* | 0.2348 | 0.2343 | 0.2398 | 0.2384 | 0.2550 | 0.2415 | 0.2343 |

**Discrepancy Index**

| Session | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|
| 1 | 0.61 | 2.27 | 2.27 | 3.40 | -6.61 | -6.61 |
| 2 | 0.00 | 2.28 | 16.89 | -13.70 | -100.00 | 68.95 |
| 3 | -0.53 | 2.35 | -0.05 | 8.96 | 0.36 | 3.22 |
| 4 | -1.61 | 0.11 | 3.82 | 5.56 | -1.27 | 7.25 |
| 5 | -3.57 | 19.64 | -4.46 | 41.96 | -100.00 | 179.46 |
| 6 | 0.30 | 0.30 | 3.87 | 5.31 | -2.18 | 20.36 |
| 7 | -2.72 | 4.51 | 5.59 | 15.78 | -51.96 | 44.74 |
| 8 | -3.50 | -3.50 | 3.39 | -37.18 | -58.19 | 21.02 |
| *Mean* | -1.38 | 3.49 | 3.91 | 3.76 | -39.98 | 42.30 |
| *SD* | 1.71 | 6.94 | 6.10 | 22.73 | 43.67 | 60.49 |

Note. PI = partial-interval sampling; WI = whole-interval sampling; MTS = momentary time sampling. Number after the hyphen indicates sampling interval in seconds.

Appendix C-2

Behavior Duration and Discrepancy Index of Sampling Estimates when Compared with Real Time Data (Celine)

**Solitary Play (code4) Duration Data**

| Session | Real time | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|---|
| 1 | 0.0680 | 0.0705 | 0.0641 | 0.0769 | 0.0962 | 0.0769 | 0.1154 |
| 2 | 0.3460 | 0.3301 | 0.3429 | 0.3718 | 0.4151 | 0.3019 | 0.3333 |
| 3 | 0.0665 | 0.0676 | 0.0649 | 0.0649 | 0.0484 | 0.0484 | 0.0968 |
| 4 | 0.2178 | 0.2133 | 0.2240 | 0.2500 | 0.2698 | 0.1746 | 0.2813 |
| 5 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 6 | 0.0389 | 0.0358 | 0.0409 | 0.0459 | 0.0152 | 0.0000 | 0.0303 |
| 7 | 0.3098 | 0.3151 | 0.3177 | 0.3264 | 0.3438 | 0.2666 | 0.3436 |
| 8 | 0.0898 | 0.0884 | 0.0945 | 0.1037 | 0.0926 | 0.0000 | 0.1071 |
| *Mean* | 0.1421 | 0.1401 | 0.1436 | 0.1549 | 0.1601 | 0.1086 | 0.1635 |
| *SD* | 0.1311 | 0.1284 | 0.1323 | 0.1404 | 0.1597 | 0.1235 | 0.1360 |

**Discrepancy Index**

| Session | | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|---|
| 1 | | 3.68 | -5.74 | 13.09 | 41.47 | 13.09 | 69.71 |
| 2 | | -4.60 | -0.90 | 7.46 | 19.97 | -12.75 | -3.67 |
| 3 | | 1.65 | -2.41 | -2.41 | -27.22 | -27.22 | 45.56 |
| 4 | | -2.07 | 2.85 | 14.78 | 23.88 | -19.83 | 29.16 |
| 5 | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 6 | | -7.97 | 5.14 | 17.99 | -60.93 | -100.00 | -22.11 |
| 7 | | 1.71 | 2.55 | 5.36 | 10.97 | -13.94 | 10.91 |
| 8 | | -1.56 | 5.23 | 15.48 | 3.12 | -100.00 | 19.27 |
| | *Mean* | -1.14 | 0.84 | 8.97 | 1.41 | -32.58 | 18.60 |
| | *SD* | 3.78 | 3.82 | 7.56 | 32.20 | 43.38 | 29.30 |

Note. PI = partial-interval sampling; WI = whole-interval sampling; MTS = momentary time sampling. Number after hyphen indicates sampling interval in seconds.

Appendix C-3

Behavior Duration and Discrepancy Index of Sampling Estimates when Compared with Real Time Data (Celine)

**Parallel Play (code5) Duration Data**

| Session | Real time | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|---|
| 1 | 0.0795 | 0.0737 | 0.0833 | 0.0833 | 0.0769 | 0.0769 | 0.0769 |
| 2 | 0.1378 | 0.1346 | 0.1410 | 0.1410 | 0.1132 | 0.1132 | 0.1111 |
| 3 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 4 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 5 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 6 | 0.0219 | 0.0205 | 0.0230 | 0.0204 | 0.0303 | 0.0303 | 0.0303 |
| 7 | 0.0036 | 0.0052 | 0.0052 | 0.0052 | 0.0156 | 0.0000 | 0.0000 |
| 8 | 0.1484 | 0.1463 | 0.1524 | 0.1646 | 0.1296 | 0.1111 | 0.1786 |
| *Mean* | 0.0489 | 0.0475 | 0.0506 | 0.0518 | 0.0731 | 0.0663 | 0.0496 |
| *SD* | 0.0640 | 0.0624 | 0.0655 | 0.0685 | 0.0498 | 0.0500 | 0.0669 |

**Discrepancy Index**

| | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|
| Session | | | | | | |
| 1 | -7.30 | 4.78 | 4.78 | -3.27 | -3.27 | -3.27 |
| 2 | -2.32 | 2.32 | 2.32 | -17.85 | -17.85 | -19.38 |
| 3 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 4 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 5 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 6 | -6.39 | 5.02 | -6.85 | 38.36 | 38.36 | 38.36 |
| 7 | 44.44 | 44.44 | 44.44 | 333.33 | -100.00 | -100.00 |
| 8 | -1.42 | 2.70 | 10.92 | -12.67 | -25.13 | 20.35 |
| *Mean* | 3.38 | 7.41 | 6.95 | 42.24 | -13.49 | -7.99 |
| *SD* | 16.84 | 15.10 | 15.96 | 118.80 | 39.60 | 40.99 |

Note. PI = partial-interval sampling; WI = whole-interval sampling; MTS = momentary time sampling. Number after hyphen indicates sampling interval in seconds.

## Appendix C-4

### Behavior Frequency and Discrepancy Index of Sampling Estimates when Compared with Real Time Data (Celine)

**Sharing and Cooperating (code2) Frequency Data**

| Session | Real time | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS -60 |
|---|---|---|---|---|---|---|---|
| 1 | 6 | 6 | 6 | 6 | 4 | 5 | 4 |
| 2 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 3 | 12 | 8 | 11 | 10 | 5 | 8 | 5 |
| 4 | 11 | 11 | 10 | 10 | 7 | 6 | 5 |
| 5 | 3 | 3 | 2 | 1 | 2 | 0 | 1 |
| 6 | 4 | 3 | 3 | 3 | 3 | 2 | 2 |
| 7 | 18 | 14 | 17 | 14 | 8 | 4 | 6 |
| 8 | 4 | 4 | 4 | 4 | 4 | 1 | 3 |
| *Mean* | 7.375 | 6.250 | 6.750 | 6.125 | 4.250 | 3.250 | 3.375 |
| *SD* | 5.755 | 4.460 | 5.498 | 4.764 | 2.375 | 2.964 | 1.923 |

**Discrepancy Index**

| Session | | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|---|
| 1 | | 0.00 | 0.00 | 0.00 | -33.33 | -16.67 | -33.33 |
| 2 | | 0.00 | 0.00 | 0.00 | 0.00 | -100.00 | 0.00 |
| 3 | | -33.33 | -8.33 | -16.67 | -58.33 | -33.33 | -58.33 |
| 4 | | 0.00 | -9.09 | -9.09 | -36.36 | -45.45 | -54.55 |
| 5 | | 0.00 | -33.33 | -66.67 | -33.33 | -100.00 | -66.67 |
| 6 | | -25.00 | -25.00 | -25.00 | -25.00 | -50.00 | -50.00 |
| 7 | | -22.22 | -5.56 | -22.22 | -55.56 | -77.78 | -66.67 |
| 8 | | 0.00 | 0.00 | 0.00 | 0.00 | -75.00 | -25.00 |
| | *Mean* | -10.07 | -10.16 | -17.46 | -30.24 | -62.28 | -44.32 |
| | *S.D* | 14.24 | 12.49 | 22.32 | 21.85 | 30.71 | 23.27 |

Note. PI = partial-interval sampling; WI = whole-interval sampling; MTS = momentary time sampling. Number after hyphen indicates sampling interval in seconds.

Appendix C-5

Behavior Frequency and Discrepancy Index of Sampling Estimates when Compared with Real Time Data (Celine)

**Solitary Play (code4) Frequency Data**

| Session | Real time | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|---|
| 1 | 3 | 2 | 2 | 2 | 2 | 2 | 2 |
| 2 | 11 | 11 | 9 | 9 | 3 | 5 | 4 |
| 3 | 4 | 4 | 3 | 3 | 4 | 1 | 1 |
| 4 | 12 | 12 | 11 | 10 | 5 | 4 | 5 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 5 | 5 | 4 | 5 | 5 | 0 | 1 |
| 7 | 17 | 10 | 14 | 8 | 6 | 6 | 4 |
| 8 | 8 | 8 | 7 | 6 | 4 | 0 | 1 |
| *Mean* | 7.5 | 6.5 | 6.25 | 5.38 | 3.625 | 2.25 | 2.25 |
| *SD* | 5.58 | 4.41 | 4.83 | 3.54 | 1.92 | 2.43 | 1.83 |

**Discrepancy Index**

| Session | | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|---|
| 1 | | -33.33 | -33.33 | -33.33 | -33.33 | -33.33 | -33.33 |
| 2 | | 0.00 | -18.18 | -18.18 | -72.73 | -54.55 | -63.64 |
| 3 | | 0.00 | -25.00 | -25.00 | 0.00 | -75.00 | -75.00 |
| 4 | | 0.00 | -8.33 | -16.67 | -58.33 | -66.67 | -58.33 |
| 5 | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 6 | | 0.00 | -20.00 | 0.00 | 0.00 | -100.00 | -80.00 |
| 7 | | -41.18 | -17.65 | -52.94 | -64.71 | -64.71 | -76.47 |
| 8 | | 0.00 | -12.50 | -25.00 | -50.00 | -100.00 | -87.50 |
| | *Mean* | -9.31 | -16.87 | -21.39 | -34.89 | -61.78 | -59.28 |
| | *S.D* | 17.37 | 10.18 | 17.37 | 31.06 | 33.40 | 29.22 |

Note. PI = partial-interval sampling; WI = whole-interval sampling; MTS = momentary time sampling. Number after hyphen indicates sampling interval in seconds.

## Appendix C-6

### Behavior Frequency and Discrepancy Index of Sampling Estimates when Compared with Real Time Data (Celine)

**Parallel Play (code5) Frequency Data**

| Session | Real time | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|---|
| 1 | 3 | 3 | 3 | 3 | 1 | 2 | 2 |
| 2 | 3 | 3 | 3 | 3 | 3 | 1 | 2 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 7 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 8 | 6 | 6 | 6 | 6 | 5 | 3 | 3 |
| *Mean* | 1.750 | 1.750 | 1.750 | 1.750 | 1.375 | 0.875 | 1.125 |
| *SD* | 2.120 | 2.120 | 2.120 | 2.120 | 1.770 | 1.130 | 1.270 |

**Discrepancy Index**

| Session | | PI-5 | WI-5 | MTS-10 | PI-30 | WI-30 | MTS-60 |
|---|---|---|---|---|---|---|---|
| 1 | | 0.00 | 0.00 | 0.00 | -66.67 | -33.33 | -33.33 |
| 2 | | 0.00 | 0.00 | 0.00 | 0.00 | -66.67 | -33.33 |
| 3 | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 4 | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 5 | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 6 | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 7 | | 0.00 | 0.00 | 0.00 | 0.00 | -100.00 | 0.00 |
| 8 | | 0.00 | 0.00 | 0.00 | -16.67 | -50.00 | -50.00 |
| | *Mean* | 0.00 | 0.00 | 0.00 | -10.42 | -31.25 | -14.58 |
| | *S.D* | 0.00 | 0.00 | 0.00 | 23.46 | 38.25 | 20.77 |

Note. PI = partial-interval sampling : WI = whole-interval sampling: MTS = momentary time sampling. Number after the hyphen indicates sampling interval in seconds.

Appendix D-1

ANOVA Summary Table: A Comparison of Effects of Sampling Interval Appropriateness on Duration Estimates of Play Behaviors

| | Ronnie | | | |
|---|---|---|---|---|
| Source | df | F | df (Error) | Error |
| Interval | 1 | 0.42 | 14 | 534.98 |
| Behavior | 2 | 136.65*** | 28 | 551.6 |
| Sampling | 2 | 14.8*** | 28 | 979.7 |
| Behavior x Interval | 2 | 2.29 | | |
| Sampling x Interval | 2 | 10.03** | | |
| Behavior x Sampling | 4 | 3.08* | 56 | 722.55 |
| Behavior x Sampling x Interval | 4 | 2.1 | | |

| | Celine | | | |
|---|---|---|---|---|
| Source | df | F | df (Error) | Error |
| Interval | 1 | 0 | 14 | 709.24 |
| Behavior | 2 | 0.24 | 28 | 949.44 |
| Sampling | 2 | 4.6* | 28 | 1796.83 |
| Behavior x Interval | 2 | 0.58 | | |
| Sampling x Interval | 2 | 4.71* | | |
| Behavior x Sampling | 4 | 1.43 | 56 | 2681.9 |
| Behavior x Sampling x Interval | 4 | 1.25 | | |

Note. MS = mean square errors. *p < .05 **p < .001 ***p < .0005

Appendix D-2

ANOVA Summary Table: A Comparison of Effects of Sampling Interval Appropriateness on Frequency Estimates of Play Behaviors

| | Randy | | | |
|---|---|---|---|---|
| Source | df | F | df (Error) | Error |
| Interval | 1 | 65.29*** | 14 | 56525 |
| Behavior | 2 | 33.42*** | 28 | 17403 |
| Sampling | 2 | 7.01** | 28 | 1082.8 |
| Behavior x Interval | 2 | 3.49* | | |
| Sampling x Interval | 2 | 5.58** | | |
| Behavior x Sampling | 4 | 5.41** | 56 | 270.98 |
| Behavior x Sampling x Interval | 4 | 1.8 | | |

| | Celine | | | |
|---|---|---|---|---|
| Source | df | F | df (Error) | Error |
| Interval | 1 | 48.75*** | 14 | 634.75 |
| Behavior | 2 | 9.21** | 28 | 878.14 |
| Sampling | 2 | 8.3** | 28 | 325.76 |
| Behavior x Interval | 2 | 1.18 | | |
| Sampling x Interval | 2 | 10.11** | | |
| Behavior x Sampling | 4 | 0.77 | 56 | 353.61 |
| Behavior x Sampling x Interval | 4 | 0.27 | | |

Note. Error = mean square errors *p < .05 **p < .005 ***p < .0005

# Appendix E

## Extra Complication with Ronnie's Data: Two Factor Interaction Between Sampling Methods and Social Behavior. and Between Sampling Interval and Type of Social Behavior

Ronnie's data showed a two factor interaction between sampling methods and type of social behavior on both dependent measures, and between sampling interval and type of social behavior on frequency measure only. Duration measures will be examined first, followed by frequency measures. As shown in Figure 4, when duration is averaged across sampling intervals, PI under-estimates duration of "sharing" by approximately 7% but over-estimates duration of "solitary play" and "parallel play" by approximately 15 and 3% respectively; WI under-estimates durations of "sharing" by approximately 25%, "solitary play" by approximately 17%, and "parallel play" by approximately 3%; in contrast MTS over-estimates durations of "sharing" by approximately 27%, "solitary play" by approximately 15%, and "parallel play" by approximately 10%. The degree of under- or over-estimation of duration varies depending on the combination of sampling method and behavior observed. Turning to frequency data (Figure 5), PI under-estimates "sharing", "solitary play", and "parallel play" by approximately 6, 34, and 36%

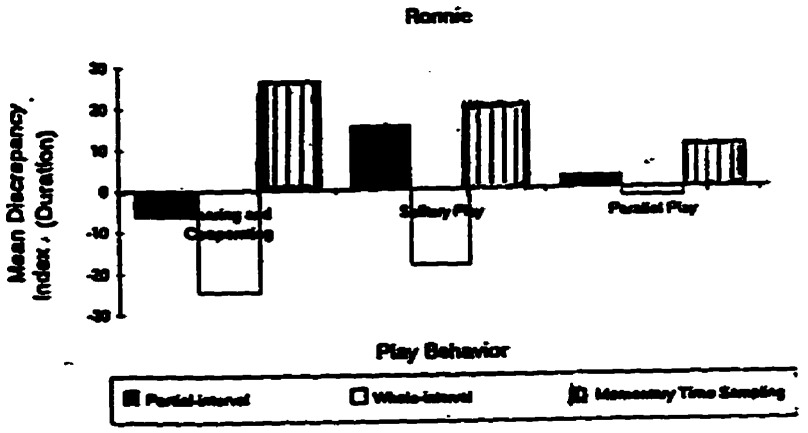



Figure 4. Duration estimates as a function of sampling methods and play behavior.

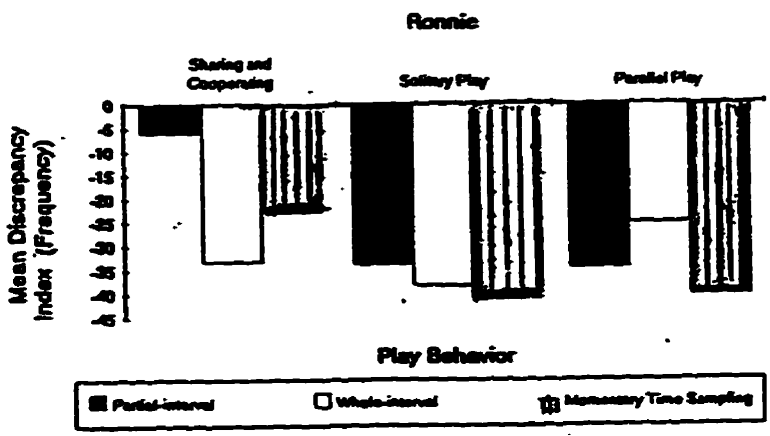



*Figure 5.* Frequency estimates as a function of sampling methods and play behavior.

respectively; WI under-estimates "sharing", "solitary play", and "parallel play" by approximately 34, 38, and 25% respectively; MTS under-estimates "sharing", "solitary play", and "parallel play" by approximately 20, 38, and 36%. Although PI, WI, and MTS all under-estimates the frequencies of the three target behaviors, PI yields the slightest under-estimation when used to score frequency data on "sharing" while the rest of sampling and behavior combinations yields large under-estimations. This interaction is found to be significant ($F[4,28] = 3.08$, $p < .02$).

Figure 6 shows the two factor interaction between sampling intervals and type of play behavior. For example, when frequency measures are averaged across sampling methods, with appropriate interval, PI under-estimates frequencies of "sharing", "solitary play", and "parallel play" by approximately 4, 12, and 18% respectively. With inappropriate interval, PI under-estimates frequency by 36, 65, and 50% respectively in the same behavioral order. This interaction is found to be significant ($F[4,28] = 5.41$, $p < .001$).

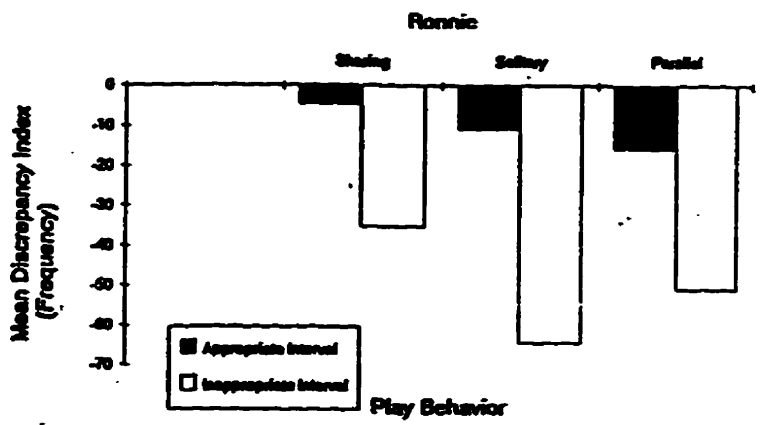



Figure 6. Frequency estimates as a function of sampling intervals and play behavior.