

A Proposed Model for Selecting Measurement Procedures for the Assessment and Treatment of Problem Behavior

Linda A. LeBlanc¹ · Paige B. Raetz¹ · Tyra P. Sellers¹ · James E. Carr²

© Association for Behavior Analysis International 2015

Abstract Practicing behavior analysts frequently assess and treat problem behavior as part of their ongoing job responsibilities. Effective measurement of problem behavior is critical to success in these activities because some measures of problem behavior provide more accurate and complete information about the behavior than others. However, not every measurement procedure is appropriate for every problem behavior and therapeutic circumstance. We summarize the most commonly used measurement procedures, describe the contexts for which they are most appropriate, and propose a clinical decision-making model for selecting measurement produces given certain features of the behavior and constraints of the therapeutic environment.

Keywords Clinical decision-making · Data collection · Measurement · Problem behavior

The systematic measurement of behavior is foundational to the delivery of applied behavior-analytic services (Baer et al. 1968; Sidman 1960). A practitioner's choices about the procedures used to track behavior over time are pivotal because direct observation data impact other important decisions. For example, proper measurement procedures allow one to examine the function of problem behavior and decide when to implement or change interventions. Indeed, there are no valid circumstances under which applied behavior analysis (ABA) should be practiced without the collection of meaningful data.

Numerous textbooks on ABA (e.g., Cooper et al. 2007; Mayer et al. 2012) and behavioral research methods (e.g., Bailey and Burch 2002; Barlow et al. 2008; Johnston and Pennypacker 2008; Kazdin 2011) describe various measurement procedures, often doing so in great detail. Some of these textbooks provide guidance about matching specific measurement procedures with specific applied circumstances (e.g., Mayer et al.). However, to our knowledge, none provide an integrated model for considering multiple aspects of an applied situation and the relative suitability of each primary measurement procedure. The recent behavioral literature for practitioners suggests one possible solution.

Clinical decision-making models provide a means to guide the selection of procedures when their optimal implementation depends on a match with specific environmental circumstances. For example, Grow et al. (2009) and Geiger et al. (2010) described decision-making models for selecting between multiple scientifically supported and function-based treatments for problem behavior maintained by attention and escape, respectively. The models consist of a series of questions that can be answered to lead a practitioner to recommendations about interventions that are optimally matched to clinical considerations (e.g., client safety, available resources, clinical goals). Fiske and Delmolino (2012) used a similar approach to present a preliminary model for selecting between discontinuous measurement procedures for problem behavior. Their model includes two important questions (i.e., initial behavior rate, terminal behavior goal) to consider when selecting between three discontinuous measurement procedures (i.e., momentary time sampling, partial-interval recording, whole-interval recording), along with recommendations for designing the selected measurement procedure. Although a useful contribution to the literature, other measurement procedures (e.g., event recording, permanent-product recording) and

✉ Linda A. LeBlanc
lleblanc@tbh.com

¹ Trumpet Behavioral Health, 390 Union Blvd., Suite 300, Lakewood, CO 80228, USA

² Behavior Analyst Certification Board, Littleton, CO, USA

practical considerations should ultimately be included in a comprehensive model of problem-behavior measurement procedure selection.

The purpose of the present article is to illustrate a clinical decision-making model for selecting between many problem-behavior measurement procedures in everyday practice. The model includes a variety of measurement procedures and practical considerations, including observability of behavior, personnel resources and constraints, dimensions of behavior, and the nature of behavior (free or restricted operant), in addition to Fiske and Delmolino's (2012) consideration about the terminal behavior goal.

Measurement Considerations and Decision-Making

Development of the Model

The model described here was developed as part of a clinical-practice standardization initiative within a large ABA human-service agency. The initiative was designed to synthesize best-practice guidelines and develop clinical decision-making tools in important areas of practice for delivering ABA services to individuals with special needs. The first three authors surveyed the published literature to identify empirical articles, literature reviews, book chapters, and textbooks on behavioral measurement. The group then synthesized the literature to develop five best-practice guidelines for collecting data on problem behavior. Implementation of the first of the guidelines—select an optimal measurement system—was facilitated by the development of the measurement summary in Table 1 and the decision-making model in Fig. 1.

Measurement Procedures

The decision-making model includes seven different measurement procedures as terminal points in the model. These procedures and their optimal circumstances are defined in the following section.

Event Recording Event recording encompasses any procedure in which the frequency of each behavior is recorded during an observation (aka frequency recording). Data from event recording are typically summarized as the frequency (count) of responses or response rate (frequency divided by time). For behaviors that have a limited opportunity to occur, such as problem behavior that only occurs in response to a task presentation, event-recording data are often summarized as the percentage of opportunities in which behavior occurred (e.g., percentage of trials with problem behavior). The primary strength of event recording is that it provides information about a specific and important behavioral dimension—its

Table 1 Features, strengths, and potential limitations of measurement procedures for problem behavior

Measurement procedure	Description	Resulting measure(s)	Strength(s)	Potential limitations
Event recording	Record each instance of behavior as it occurs	Frequency, rate, percentage of opportunities (for restricted operants)	Direct measures of the behavior	Requires constant vigilance; may be impractical for high-frequency or non-discrete behavior
Duration recording	Record the amount of time from onset to offset for each behavior as it occurs	Total duration, mean duration, percentage duration	Direct measures of the behavior (including frequency)	Requires constant vigilance; requires a timing device
Latency recording	Record the amount of time that passes between the discriminative stimulus and the onset of the behavior	Mean latency	Direct measures of the temporal relation between the discriminative stimulus and the behavior	Requires constant vigilance; requires a timing device
Intensity recording	Record a dimension of intensity (e.g., force, volume) for each instance of the behavior	Various (e.g., mean decibel, mean rating per event)	Direct measures of the behavior; automated recording possible with some dimensions (e.g., volume)	Requires a reliable and valid measurement device (e.g., decibel meter) or rating system; requires constant vigilance
Permanent-product recording	Document the effects of a behavior on the environment	Various (e.g., wound size, number of holes in wall)	Allows measurement of behavior that occurs at inaccessible times	An indirect assessment of behavior; behavior must reliably produce the product and be the only source of the product
Partial-interval recording	Record whether a behavior occurred at all during specific time intervals for defined observation period(s)	Percentage of intervals in which the behavior occurred	Does not require constant vigilance	Generates an estimate of behavior; systematically overestimates the occurrence of behavior; requires a timing device
Momentary time sampling	Record whether a behavior occurred at a given moment for defined observation period(s)	Percentage of samples or intervals in which the behavior occurred	Allows concurrent measurement of multiple individuals or behaviors; does not require constant vigilance; good correspondence to event recording compared to other discontinuous procedures	Generates an estimate of behavior. Inappropriate for short-duration or low-frequency behavior; requires a timing device.

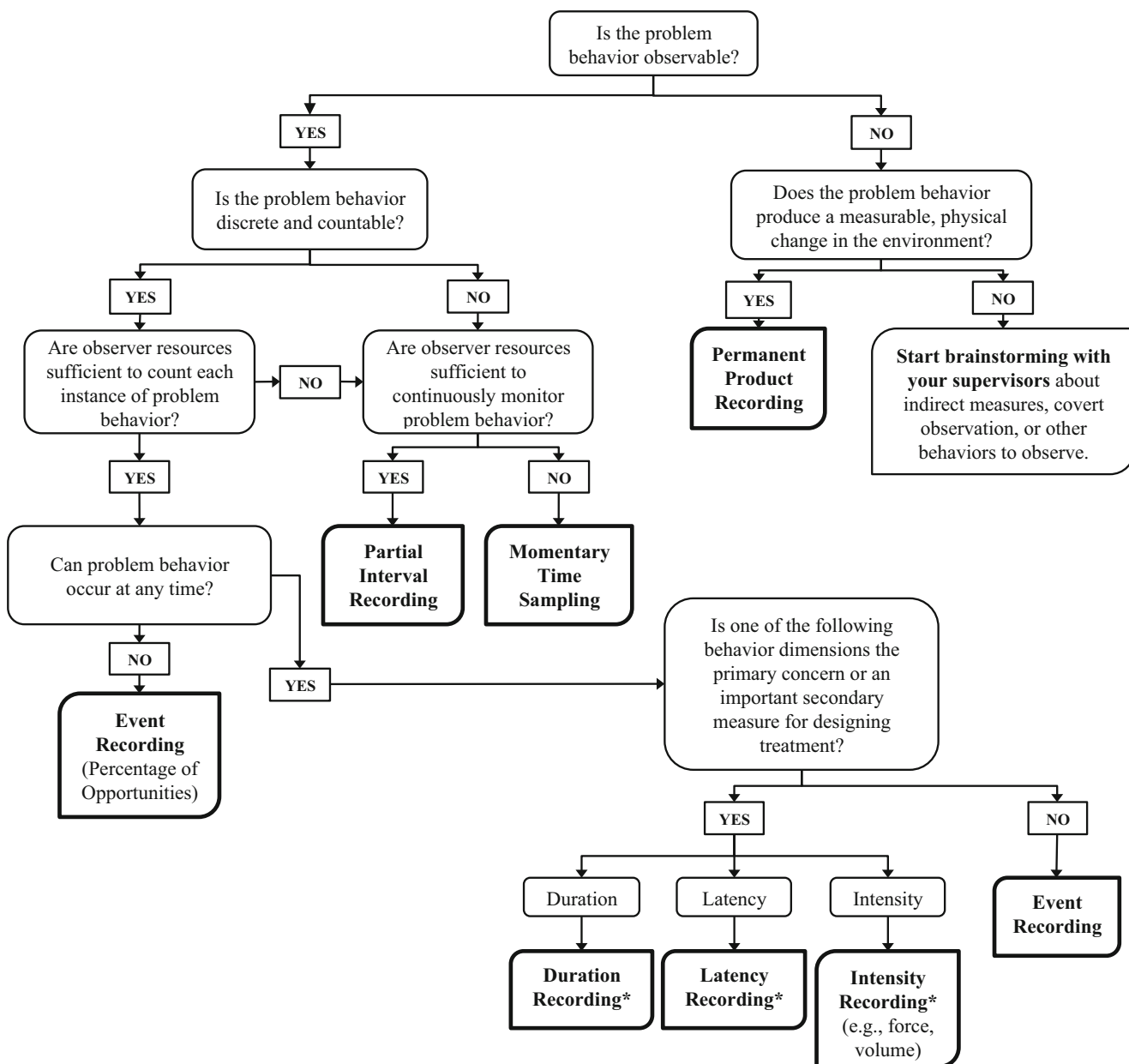


Fig. 1 A decision-making model for selecting measurement procedures for problem behavior. Note: an asterisk also generates a frequency count

frequency of occurrence. However, event recording is only appropriate for behaviors that have clear beginnings and endings and do not occur so frequently that it is impossible to accurately record them. Event recording is also best suited for behaviors that occur for comparable durations. In addition, event recording requires the observer to constantly monitor the behavior, which may not be feasible in many service environments.

Duration Recording Duration recording involves the measurement of the amount of time each behavior occurs during an observation. Duration-recording data are typically summarized as the mean duration of each behavior during

an observation or the total duration of all behaviors during an observation. The latter measure might also be expressed as the proportion of the observation in which behavior occurred. Duration recording is a preferred measurement procedure when information about how long the behavior occurs is the most relevant dimension of interest. In addition, duration recording in which the duration of each behavior is documented also generates a frequency measure. Like event recording, however, duration recording requires constant vigilance, which might limit its practicality. Duration recording also requires a timing device (e.g., stopwatch, clock app) that must be easily accessible yet unobtrusive.

Latency Recording Latency recording involves the measurement of the amount of time (usually in seconds) it takes each behavior to occur following a specific environmental event (e.g., a discriminative stimulus) during an observation. Latency-recording data are typically summarized as the mean latency to each behavior during an observation. Latency recording is a preferred measurement procedure when information about a behavior's latency is the dimension of interest. For example, Call et al. (2009) evaluated the mean latency to problem behavior during different tasks as an index of each task's aversive properties (i.e., tasks associated with low latencies were aversive). In addition, latency recording also generates a frequency measure. However, like duration recording, latency recording requires constant vigilance (for both the behavior and antecedent event) and a timing device.

Intensity Recording Intensity recording involves the measurement of the intensity (or magnitude) of each behavior during an observation. Intensity might be recorded by the behavior's force, loudness, or other relevant characteristic. Recording these characteristics can be done objectively (e.g., decibels) or subjectively (e.g., rating a behavior's force on a 5-point scale). How intensity-recording data are summarized after an observation depends on the specific measured characteristic, but the mean intensity is likely a relevant measure for many situations. Alternatively, the percentage of behaviors in an observation that exceeded a certain threshold might also be relevant. Intensity recording is a preferred measurement procedure when information about a behavior's intensity is the dimension of interest. In addition, intensity recording also generates a frequency measure. However, intensity recording requires constant vigilance, as well as a reliable and valid measurement system. The latter requirements are especially important when intensity recording involves equipment or a subjective rating system.

Permanent-Product Recording Permanent-product recording involves the measurement of behavior by its physical impact on the environment. Examples of permanent-product recording include measuring tissue damage from self-injury (e.g., Grace et al. 1996) and counting the number of food items missing as evidence of food stealing (e.g., Maglieri et al. 2000). How permanent-product recording data are summarized after an observation depends on the specific evidence produced by the behavior, but the mean or frequency of these events is likely a relevant measure for many situations. Permanent-product recording is a preferred measurement procedure when direct observation of the behavior is impossible or impractical. However, the utility of this procedure is limited by its indirect nature (i.e., behavior is not directly observed) and by the following required conditions: the behavior must reliably produce the product, and the product must not be frequently produced by any other behavior or event.

Partial-Interval Recording Partial-interval recording involves documenting whether the behavior occurs in each of a consecutive series of brief time periods. Partial-interval data are typically summarized as the percentage of intervals in which behavior was scored. Because partial-interval recording does not require constant vigilance (i.e., one no longer needs to observe within an interval after a behavior has been scored), it is often used to measure high-rate behavior, as well as multiple forms of behavior. However, partial-interval recording is associated with several disadvantages. Unlike event, duration, latency, and intensity recording, which are continuous recording procedures, partial-interval recording is a discontinuous measurement procedure because some behaviors are deliberately not recorded. Thus, partial-interval recording does not produce complete data about any behavioral dimension but instead generates an estimate of the frequency and duration of behavior. In addition, partial-interval recording consistently overestimates the level of behavior, and this overestimation is exaggerated by long intervals and observation periods (Wirth et al. 2013). We refer the reader to Fiske and Delmolino (2012) for a discussion about the determination of interval and observation-period sizes.

Momentary Time Sampling Momentary time sampling involves documenting whether the behavior occurs at the end of each of a consecutive series of brief time periods. Momentary time sampling data are typically summarized as the percentage of intervals in which behavior was scored. Because momentary time sampling does not require constant vigilance, it is often used to measure multiple forms of behavior or the behavior of multiple individuals. However, momentary time sampling is associated with several disadvantages. Momentary time sampling is a discontinuous measurement procedure and thus does not produce complete data about any behavioral dimension but instead generates an estimate of behavior. In addition, the type of error generated by momentary time sampling is inconsistent (unlike the overestimation of partial-interval recording) and is exaggerated by long intervals and observation periods (Wirth et al. 2013). We refer the reader to Fiske and Delmolino (2012) for a discussion about the determination of interval and observation-period sizes.

Using the Model

The decision-making model consists of a series of questions that a practitioner can ask and answer to select the optimal measurement procedure for problem behavior given specific client and environment circumstances. This model is intended for use by behavior analysts with experience in the assessment and treatment of problem behavior and measurement. The model depicted in Fig. 1 may be most useful to those who are relatively new to practice and in need of a systematic guide

to selecting optimal measurement procedures for problem behavior.

The questions in the model focus on several important considerations in the order that logically governs selection of a measure: (a) specific characteristics of the behavior, (b) personnel resources and constraints, (c) important dimensions of behavior, and (d) the nature of behavior as a free or restricted operant. That is, behaviors that occur covertly, or cannot be directly observed, present unique measurement concerns that often prevent direct observation, rendering all of the other questions moot. Thus, this question is posed first in the manuscript and a “yes” answer ends the process with options to use permanent-product recording or to brainstorm other strategies for gaining access to the behavior in real time (i.e., overtly). The next question refers to whether the behavior is discrete and countable, regardless of resources, in order to make the critical decision between a continuous (e.g., event, duration, latency, and intensity recording) and a discontinuous (e.g., time sampling, interval recording) measurement procedure. Some behaviors may be initiated as a single instance but occur for a long duration (e.g., vocal stereotypy, hand mouthing, off-task behavior), making it difficult for a continuous measurement procedure to capture and provide the most useful information about the behavior. When it is possible to capture information on each event, continuous measurement procedures are strongly preferred because they eliminate the error associated with estimate-based discontinuous measurement (see prior descriptions of momentary time sampling and partial-interval recording). Some discontinuous measurement procedures (i.e., whole-interval recording) tend to systematically underestimate the actual level of behavior and thus are inappropriate for measuring problem behavior. For this reason, whole-interval recording is not recommended in our model.

Personnel and resource constraints figure into the model next. Note that even if the behavior itself is discrete and observable and thus, amenable to continuous measurement, the available resources may constrain the use of such procedures. If a person cannot remain constantly vigilant for all instances of behavior (e.g., a teacher in a classroom with 25 students), then a discontinuous measurement procedure in the form of momentary time sampling would likely be a more practical and preferred option. If there are sufficient resources to collect data using a continuous measurement system, then the next questions lead a practitioner to select the optimal measurement procedure given the nature of the behavior as a free or restricted operant and the important dimensions of the behavior.

Many behaviors can occur at any time and in any setting (i.e., a free operant) while others can only occur under certain circumstances (i.e., a restricted operant). That is, with a restricted operant behavior, there must be some condition in place for the behavior to occur. For example, some instructional event must be presented and occur in an ongoing

manner for off-task behavior to occur. Off-task behavior would likely be recorded using duration recording unless resources prohibited constant vigilance, in which case a discontinuous measurement procedure (e.g., momentary time sampling) would be used. As another example, a demand must be presented in a context in which compliance could reasonably occur in order for there to be any opportunity for noncompliant behavior to occur. Noncompliance would likely be scored as the percentage of opportunities or demands resulting in noncompliant behavior. It may also be the case that protective equipment or the absence of some person or material eliminates the possibility of the behavior occurring. For example, protective arm splints might be used to prevent the occurrence of very severe hand or arm biting or intense head hitting. As another example, aggression towards a sibling can only occur when the sibling is present. In these instances, it is best to collect a continuous measure (e.g., frequency, duration) but use a denominator that reflects the restricted observation window in which the behavior can occur in the conversion to a rate or percentage duration measure.

If the behavior can occur at any time, consider all dimensions of the response and select the ones that are most critically important to fully capture the important features of the behavior and the potential change in the behavior that may occur due to intervention. This should lead you to select the most appropriate type of continuous measurement procedure or a combination of those procedures. The most commonly used continuous recording procedure is event recording in which a frequency (count) is often converted to a rate measure by dividing it by the duration of the observation window. The inter-response time (IRT) can also be calculated from event recording to inform intervention planning (e.g., the start interval for NCR). When other dimensions of the behavior (e.g., duration, intensity) are not particularly important or are relatively equal across instances of the behavior, event recording is the recommended procedure. Note that it is not particularly useful to have a frequency count without making note of the observation time in order to subsequently convert it to a rate. Even when it seems like an observation window will always be the same (e.g., the school day), certain factors may change the duration of the observation at some point (e.g., the child leaves school ill, an unexpected field trip occurs).

When some other dimension of behavior can vary greatly, capture a measure of that dimension (e.g., duration, intensity) and derive the frequency count and rate from the number of observed instances. That is, each event is scored for intensity, so the number of scored events is equal to the frequency count. Tantrums are often scored using a duration measure but may also include an intensity rating or a frequency count of a specific intense problem behavior that is then divided by the duration of the tantrum (e.g., rate of aggression). Over the course of an intervention, one might observe a change in the number of tantrums, the average duration of each tantrum, or

the rate of aggression during tantrums. Other important and useful measurement procedures, such as latency recording, not only produce information about the specific temporal dimension, a frequency measure can be obtained from the data and in many instances a rate measure as well.

Case Example

Joey is a 7-year-old student in a classroom with 22 other students. The teacher and paraprofessional aide are willing to collect data for the behavior analyst who is consulting on the case. Joey is often off-task and disruptive during independent seatwork, and he says rude things to other students in the class, as well as the adults. The behavior analyst used the clinical decision-making model to select measurement procedures for the teacher and aide to use. The teacher and aide indicated that the rude statements occur multiple times per day and are discrete, noticeable, and often reported by other students. As long as the teacher and aide do not have to record exactly what Joey says when he misbehaves, they can collect data on each event that occurs almost immediately after it occurs. The behavior is not covert (i.e., it is observable), there are sufficient resources for continuous measurement, the behavior could occur at any time, and there is no other critically important dimension of behavior besides occurrence. Thus, the model leads the behavior analyst to use event recording and calculate a frequency per school hour.

During independent seatwork, both the teacher and the aide are roaming the classroom providing assistance to various students and cannot remain vigilant to continuously score Joey's attending or inappropriate behavior. They often have difficulty telling whether he is actively engaged in his work or is looking down towards his desk while thinking about other things. They are both in the room and can observe periodically but not continuously, as long as they do not have to be right next to him when the observation occurs. There is at least one component of his behavior that is somewhat covert as they cannot always tell whether he is actively engaged in the work at any given moment from across the room. Thus, there could be a permanent product measure from each observation such as the percentage of the assigned work completed. This is not a direct measure of on-task behavior, but it can be a useful supplement to other measures. Disruptive behaviors may be discrete and countable, but other off-task behaviors are not, and the environment does not have sufficient resources for continuous observation. Thus, the permanent product measure could be supplemented with momentary time sampling with either the teacher or aide periodically (e.g., every 3 or 5 min) recording whether Joey appears to be writing or reading or otherwise working on his task and whether any disruptive behavior occurs at the observation point. These data would then be converted to a percentage-of-observations measure.

Conclusion

Practicing behavior analysts frequently assess and treat problem behavior as part of their ongoing employment responsibilities. Effective measurement of problem behavior is critical to success in these activities because some measures of problem behavior provide more accurate and complete information about the behavior than others. However, not every measurement procedure is appropriate for every problem behavior. In addition, the resources available on an ongoing basis in natural environments may not always support the most labor-intensive measurement procedures. One concern is that behavior analysts who encounter barriers to complicated or optimal data collection systems might fail to collect data altogether if they do not have a system for selecting the most useful procedures given their constraints. Another concern is that one might select a measurement procedure that does not provide sufficient information about the behavior to allow a meaningful evaluation of the effects of a given intervention.

Other clinical decision-making models have been developed to guide behavior analysts in selecting among options for intervention for problem behavior when multiple function-based treatments have evidence to support their potential effectiveness (e.g., Geiger et al. 2010; Grow et al. 2009). The use of this type of model may introduce a comprehensive and thoughtful framework for decision-making when choices might otherwise be guided by familiarity with only a few of the options. Similarly, selections of measurement procedures for problem behavior might also be determined by recent use or prior history of a procedure not working well in a different context. Fiske and Delmolino (2012) provided an example of how to consider the pros and cons for a more limited set of procedures (i.e., discontinuous measurement procedures only), and the current model expands this idea to a broader range of potential measurement procedures for problem behavior.

This article may provide multiple benefits to the applied user. First, the table and text describing each measure provides a convenient and succinct summary of the strengths and considerations for the most frequently used measurement procedures for problem behavior. Second, the selection model may guide practitioners through the most commonly encountered barriers to effective data collection: (a) inability to observe the behavior, (b) lack of resources for continuous data collection, and (c) a mismatch between the properties of the behavior itself and the procedure. When there are no specific barriers and there are no other specific dimensions of behavior that require special attention, the resulting selection is a frequency count that can be converted into a rate measure. When barriers exist or it is important to capture other dimensions of behavior, the resulting selections are the measurement procedures best suited to those circumstances. Additionally, if a practitioner has already unsuccessfully attempted to use one measurement procedure that may have been optimal from a technical

perspective, this model might assist them in selecting the best of the remaining possible procedures that is better suited for their constraints.

We hope that practicing behavior analysts will find the article useful when they are selecting measures for a problem behavior or environmental constraint that they have not addressed before or for refining their existing experience and expertise in selecting measurement procedures for problem behavior. We also hope that this article and the others like it (e.g., Geiger et al. 2010; Grow et al. 2009) might assist practitioners in using a more systematic approach in all of their critical clinically related decisions and processes (e.g., measurement and data collection, treatment planning, curricular assessment).

Finally, the model presented herein has not been empirically tested to determine if measurement procedures selected based on the decision-making model produce more comprehensive or sensitive evaluations of the effects of interventions than measures selected without the use of the model. Thus, there is no empirical demonstration of any differential effects of this model on the ease or utility of practitioners' measurement efforts. However, because the model is based on a well-developed literature on measurement, it might be prudent to consider it a potentially useful starting point until validation data can be generated. Towards that end, research efforts might include assessments of the model's social validity (i.e., are the selected measures easier to use and do they result in more generated data) and empirical comparisons between model-generated and default measurement procedures in terms of improved data-based decision-making.

Author Note This article does not represent an official position of the Behavior Analyst Certification Board. Paige Raetz is now at the Southwest Autism Research and Resource Center. Tyra Sellers is now at Utah State University. This model was developed as part of the Clinical Standards initiative at Trumpet Behavioral Health.

References

- Baer, D. M., Wolf, M. M., & Risley, T. (1968). Current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*, 91–97.
- Bailey, J. S., & Burch, M. R. (2002). *Research methods in applied behavior analysis*. Thousand Oaks: Sage.
- Barlow, D. H., Nock, M. K., & Hersen, H. (2008). *Single case experimental designs: strategies for studying behavior change* (3rd ed.). Boston: Allyn and Bacon.
- Call, N. A., Pabico, R. S., & Lomas, J. E. (2009). Use of latency to problem behavior to evaluate demands for inclusion in functional analyses. *Journal of Applied Behavior Analysis, 42*, 723–728.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River: Pearson.
- Fiske, K., & Delmolino, L. (2012). Use of discontinuous methods of data collection in behavioral intervention: guidelines for practitioners. *Behavior Analysis in Practice, 5*(2), 77–81.
- Geiger, K. A., Carr, J. E., & LeBlanc, L. A. (2010). Function-based treatments for escape-maintained problem behavior: a treatment selection model for practicing behavior analysts. *Behavior Analysis in Practice, 3*(1), 22–32.
- Grace, N. C., Thompson, R., & Fisher, W. W. (1996). The treatment of covert self-injury through contingencies on response products. *Journal of Applied Behavior Analysis, 29*, 239–242.
- Grow, L. L., Carr, J. E., & LeBlanc, L. A. (2009). Treatments for attention-maintained problem behavior: empirical support and clinical recommendations. *Journal of Evidence-Based Practices for Schools, 10*, 70–92.
- Johnston, J. M., & Pennypacker, H. S. (2008). *Strategies and tactics of behavioral research* (3rd ed.). New York: Routledge.
- Kazdin, A. E. (2011). *Single-case research designs: methods for clinical and applied settings* (2nd ed.). New York: Oxford University Press.
- Maglieri, K. A., DeLeon, I. G., Rodriguez-Catter, V., & Sevin, B. M. (2000). Treatment of covert food stealing in an individual with Prader-Willi syndrome. *Journal of Applied Behavior Analysis, 33*, 615–618.
- Mayer, G. R., Sulzer-Azaroff, B., & Wallace, M. (2012). *Behavior analysis for lasting change* (2nd ed.). Cornwall-on-Hudson: Sloan.
- Sidman, M. (1960). *Tactics of scientific research: evaluating experimental data in psychology*. New York: Basic Books.
- Wirth, O., Slaven, J., & Taylor, M. A. (2013). Interval sampling methods and measurement error: a computer simulation. *Journal of Applied Behavior Analysis, 47*, 83–100.