Improving Accuracy of Data Collection on a Psychiatric Unit for Children Diagnosed With Intellectual and Developmental Disabilities

Patrick W. Romani^{1,2} · Aimee S. Alcorn³ · James Linares³

Published online: 1 November 2018
© Association for Behavior Analysis International 2018

Abstract

Data collection is a hallmark of effective behavior-analytic therapy. Collecting accurate data permits a behavior analyst to evaluate the effectiveness of behavioral treatment. The current study evaluated the use of a clicker, simplified observation, and a timer to improve accuracy of data collection on a psychiatric unit for children diagnosed with intellectual and developmental disabilities. Experiment 1, conducted within a combined multiple-baseline across-participants and reversal design, was an evaluation to identify an intervention package for four participants employed by the psychiatric unit. Interventions yielding the highest interobserver agreement (IOA) were highly individualized. Thus, we selected the most comprehensive intervention and exposed four additional participants to this intervention during Experiment 2. Results showed that this intervention improved IOA for these additional participants as evaluated within a multiple-baseline across-participants design. Results of the current study will be discussed to assist other behavior analysts in improving data-collection practices in hospital or school settings.

Keywords Data collection · Interobserver agreement · Staff training · Psychiatric unit

Applied behavior-analytic work assesses individual behavioral excesses or deficits before evaluating behavior change following manipulation of an independent variable (Asmus, Vollmer, & Borrero, 2002; Tiger, Hanley, & Bruzek, 2008). To study behavior change, data need to be collected on either the frequency or the duration of the target behavior. Only then can objective evidence show the effects of a treatment program.

There is a long history of discussing the importance of data-collection procedures in the field of applied behavior analysis (Wolf, 1978). In fact, Baer, Wolf, and Risley (1968) described precise measurement as one way to meet the behavioral expectation of the field. That is, quantification of behavior must be in a reliable manner so that behavior change can be objectively documented. Without such reliable quantification,

a “believable demonstration of the events that can be responsible for the occurrence or non-occurrence of that behavior” cannot occur (Baer et al., 1968, pp. 94–95). Moreover, data are necessary as they can impact potentially life-changing intervention (Vollmer, Sloman, & St. Peter Pipkin, 2008).

It is likely that all behavior analysts recognize the importance of using accurate data to guide treatment decisions. However, practitioners may find it difficult to introduce data-collection procedures that maintain with sufficient integrity in their absence (Sigurdsson & Austin, 2006; Wooderson, Cuskelly, & Meyer, 2014). For example, if a behavior analyst spends 1 h conducting an initial evaluation of treatment procedures, it would likely be helpful for maintenance to be evaluated via continued data collection. Fradenburg, Harrison, and Baer (1995) discussed characteristics of the observers (e.g., training history), the nature of the data-collection system (e.g., number of behaviors to score), and the data-collection setting (e.g., number of children, resources) as three factors that may interfere with accurate data collection. The independent, and often combined, influence of these variables may be important for behavior analysts to consider when establishing data-collection programs.

Blough et al. (2006) used training and supervisor feedback to increase data-collection behaviors for direct-care staff collecting data on client response to psychotropic medication. The treatment package increased frequency of data collection

✉ Patrick W. Romani
Patrick.romani@childrenscolorado.org

¹ School of Medicine, University of Colorado, Anschutz Medical Campus, Aurora, CO 80045, USA

² Children’s Hospital Colorado, 13123 E. 16th Ave, Aurora, CO 80045, USA

³ Department of Psychiatry, Children’s Hospital Colorado, 13123 E 16th Ave, Aurora, CO 80045, USA

for these direct-care staff. It was not clear that the data collected and graphed by the direct-care staff were accurate, though. In an extension, Mozingo, Smith, Riordan, Reiss, and Bailey (2006) addressed data-collection accuracy. These researchers employed a combination of supervisor feedback and supervisor presence to improve the accuracy of frequency-data collection for direct-care staff in a residential treatment setting monitoring problem behaviors. Results showed increased accuracy with this treatment package.

These studies addressed characteristics of the observer to establish data-collection procedures (Blough et al., 2006; Mozingo et al., 2006) and characteristics of the data-collection setting (e.g., supervisor observation; Mowery, Miltenberger, & Weil, 2010; Mozingo et al., 2006). Fewer studies have addressed the nature of the data-collection system. In one exception, Mash and McElwee (1974) found improved accuracy when observers collected data on fewer behaviors. In most settings, though, it is likely that a combination of these strategies is needed to improve data-collection accuracy. Studies need to show ways to account for all three of these threats to data-collection accuracy to improve data-collection practices in applied settings.

The purpose of the current study was to evaluate the components of a data-collection program to increase the accuracy of data collection, as measured through interobserver agreement (IOA). We evaluated training (characteristics of the observers), modifications to the observation system to make operational definitions more clear (nature of the data-collection system), and the addition of clickers and signals (characteristics of the setting) as interventions to improve the accuracy of data collection. During Experiment 1, we evaluated which of these interventions would improve the quality of data collection. Once we determined the conditions under which all four participants collected data accurately, we exposed four additional participants to this program to establish the validity of these data as a way to improve accuracy of data collection across the broader psychiatric unit (Experiment 2).

Experiment 1

Method

Participants

Direct-Care Staff (Hereafter Referred to as “Participants”) Four participants, employed by a specialized psychiatric inpatient and partial hospitalization unit providing multidisciplinary treatment to children diagnosed with intellectual and developmental disabilities, participated in this study. Participants were responsible for simultaneously collecting data and implementing behavioral treatment plans. Three of the participants had a bachelor’s degree in psychology or a related field and one participant had a master’s degree in counseling. All

participants consented to be part of the study and were assured that their job performance would not be evaluated based on participation in this study. Participants received a 60-min didactic presentation on data-collection practices as part of initial training to work on the psychiatric unit. Modeling of frequency-based data collection, practice using video exchanges between confederates, and immediate feedback occurred during the didactic presentation.

Patients The participants’ collection of behavioral data was observed as they worked with patients admitted to the psychiatric unit. Please see Table 1 for specific information about presenting patient diagnoses and problem behaviors.

Advanced Data Coders Advanced data coders were the first two authors. Each had received training in behavioral data coding as part of their course sequence to become Board Certified Behavior Analysts. Additionally, they both had collected data with one another and showed a high level of agreement (i.e., above 80% agreement) within the context of the current study.

Setting and Materials Observations were conducted in classrooms on the psychiatric unit. Classrooms were approximately 4.5 m × 6.1 m. Each classroom was equipped with six desks, one large table, and an array of leisure activities.

Initially, the participants only had a data-collection sheet (20.32 cm × 25.4 cm) on which they were instructed to take data on multiple target problem and appropriate behaviors during their 8- or 12-h shifts. The data-collection sheet was divided into 15-min intervals. Participants were expected to write the number of problem and appropriate behaviors that occurred during each 15-min interval. For participants that needed the support of a clicker, the clicker fit on the participants’ name badges. Participants were instructed to touch a button on the clicker to record each instance of problem and appropriate behavior. A timer was used for Participant 4, in which a sound was produced when the timer expired.

Response Definitions, Observation System, and IOA The dependent variable for the current study was data coding. Data coding was defined as the participant writing down the correct number of behaviors observed during each 15-min interval in the correct interval. For example, if four instances of aggression occurred from 9:00 a.m. to 9:15 a.m., the participant should have written the number 4 within the “9:00 a.m.” interval on the data sheet.

The advanced data coders observed the patients that the four participants were assigned to collect data on from a discrete location in the classroom on the psychiatric unit. The advanced data coders collected data along with the participant during each 15-min interval. These data were later compared to produce IOA. Total count IOA was used to calculate IOA between the advanced data coders and the participants. That

Table 1 Demographic information for patients

Patient	Age (years)	Diagnoses	Target Behavior	Experiment
001	17	ASD; unspecified schizophrenia spectrum disorder	Self-talk	1 only
002	8	ASD; unspecified anxiety disorder	Aggression and property destruction	1 only
003	17	ASD; bipolar disorder	Screaming	1 and 2
004	15	ASD	Aggression, self-injury, and property destruction	1 only
005	17	ASD; moderate ID	Aggression and self-injury	1 and 2
006	15	Mild ID	Inappropriate vocalizations	1 and 2
007	13	ASD; unspecified anxiety disorder	Aggression	1 only
008	15	ASD; moderate ID	Property destruction	2 only
009	6	ASD; unspecified anxiety disorder	Property destruction	2 only
010	8	ASD; moderate ID	Property destruction	1 and 2
011	12	ASD; CP; moderate ID	Self-injury	2 only

ASD Autism spectrum disorder, ID Intellectual disability, CP Cerebral palsy

is, the number of agreements were divided by the number of agreements plus disagreements and multiplied by 100 to produce a total count IOA percentage for that 15-min interval. If a participant transitioned to a condition that included simplified observation, the advanced data coder collected data on all target behaviors and recorded those on the data sheet for the participant so as not to affect overall analysis of patient data.

On an average of 29.8% of intervals, the second advanced data coder collected data alongside the first advanced data coder. Total count IOA was calculated according to the same procedure described earlier. For Participant 1, IOA was collected on 28.6% of intervals and was 100%. Total count IOA for Participant 2 was collected on 33.3% of intervals and was 100%. Total count IOA for Participant 3 was collected on 31.8% of intervals and was 100%. Total count IOA for Participant 4 was collected on 30.4% of intervals and was 100%.

Experimental Design A combined multiple-baseline across-participants and reversal design was used. The purpose of the multiple-baseline across-participants design was to show functional control over the interventions to improve data collection. The reversal design was used to show functional control over the intervention that produced increased IOA for each individual participant.

Procedure Advanced data coders randomly selected one or two 15-min intervals to observe each participant per week. Participants were observed while collecting data for multiple patients. Before each observation interval, the participant was told to use the clicker, review the simplified observation, or respond to the signal. They were directed to follow typical data-collection practices in between these observations. Conditions for the current study were conducted until individual participant IOA improved to above 80% for at least two consecutive 15-min intervals.

Baseline (Condition A) During Condition A, participants were directed to collect data on the dependent variables for the patient they were working with that day. A behavioral treatment plan was provided to the participants that contained operational definitions for the patient's specific behaviors they were tracking. Participants needed to track multiple target behaviors for decrease (e.g., aggression, property destruction, self-injury) and one target behavior for increase (e.g., communication, compliance with instructions).

Clicker Only (Condition B) During Condition B, the participants were again provided with the patient's behavioral treatment plan and were told to review the patient's operational definitions. Participants were then provided with a clicker to track the behaviors for decrease and increase. The use of the clicker was described to the participants so they understood how and when they were supposed to use the clicker.

Clicker Plus Simplified Observation (Condition C) During Condition C, the participants were presented with the patient's behavioral treatment plan. They were also directed to review the patient's operational definitions. The number of definitions within the behavioral treatment plan was simplified so that the participant was only taking data on the most frequently occurring topography of behavior for decrease. For example, if a patient was observed to most commonly engage in hand biting but also on occasion engaged in head banging, the participant was directed to only attend to the operational definition for hand biting. Participants were then provided clickers to track data on this topography of problem behavior.

Clicker, Simplified Observation, and Signal (Condition D) One participant participated in this condition of the current study. During Condition D, the procedures described in Condition C were followed. When the advanced data coder observed the

participant taking data, a timer was set for 15 min. Before the observation began, the participant was told to record data using the clicker during the interval and to record the frequency of behavior occurrence on the data sheet when the participant heard the timer sound.

Results and Discussion

Figure 1 describes responding for Participant 1 (top panel), Participant 2 (second panel), Participant 3 (third panel), and Participant 4 (bottom panel). During the initial baseline condition (Condition A), average IOA across all participants was 54.96% (range 0%–100%). Participant 1 showed 100% IOA across seven baseline 15-min intervals. Participant 2's IOA during baseline was initially in the moderate range before decreasing. Average IOA was 35% (range 20%–50%) for this participant. Participant 3 initially collected data accurately before IOA values

decreased to zero or near-zero levels for three consecutive 15-min intervals. Average IOA for Participant 3 was 35.7% (range 0%–100%). Participant 4's responding was highly variable and averaged 35% (range 0%–80%).

After three of four participants showed a performance deficit, the clickers were introduced (i.e., Condition B) for Participants 2, 3, and 4. Participant 2 responded positively to the clickers. Participant 2's IOA increased and averaged 89.8% (range 59.4%–100%). In contrast, Participants 3 and 4 did not respond positively to Condition B. Participant 3's IOA initially increased to 100% for the first two 15-min intervals but then decreased to baseline levels for the final three 15-min intervals. Average IOA during this condition was 53.5% (range 0%–100%). Participant 4's responding seemed to be consistently within the moderate range until the fifth 15-min interval of this condition, which decreased to 0%. The variability evident in this participant's baseline condition persisted; average IOA was 56.7% (range 0%–100%).

Fig. 1 Total IOA for Participants 1–4 during Experiment 1. Open data points indicate no target behaviors were observed

Participant 2 reversed back to baseline after showing positive response to Condition B. IOA values decreased to baseline levels for Participant 2. Average IOA for Participant 2 was 29.2% (range 25%–33.3%) during this return to baseline.

Participants 3 and 4 transitioned to Condition C. Participant 3 showed an increasing trend in IOA when exposed to this condition; IOA averaged 84.9% (range 58.8%–95.8%). Unfortunately, Participant 4 did not respond positively to this intervention; IOA continued to be exhibited at variable levels ($M = 33.3\%$; range 0%–100%).

Participant 2 then transitioned back to Condition B after reestablishing baseline levels of responding without this intervention. This participant responded positively to the reimplementation of this intervention. Participant 2's IOA increased and maintained at high levels ($M = 96\%$; range 80%–100%). Participant 3 also transitioned back to Condition B; IOA decreased to baseline levels ($M = 14.3\%$; range 0%–42.9%). Participant 4 was transitioned into Condition D (clicker, simplified observation, and signal condition); IOA values increased to 100% and remained stable for three consecutive 15-min intervals.

Participant 3 transitioned back to Condition C after reestablishing baseline levels of IOA. After one 15-min interval with low IOA, IOA increased and stabilized at 100% for three consecutive 15-min intervals ($M = 80\%$; range 20%–100%). Participant 4 transitioned back to this condition after showing a positive response to Condition D, the whole treatment package. IOA values immediately decreased to baseline levels ($M = 25\%$; range 0%–50%). Participant 4 then transitioned back to Condition D; IOA values increased and stabilized at 100% for two consecutive 15-min intervals. For Participants 2–4, several intervals that were observed contained high-rate behavior, as defined by 10 or more instances of a target behavior. During these periods, IOA was low ($M = 36.4\%$; range 15%–56%).

In summary, each participant showed that a different level of intervention was needed to improve IOA. One participant responded well to baseline data-collection procedures. A second participant's IOA improved to sufficient levels following the addition of a clicker to record data during each 15-min interval. One participant benefited from simplified observations, in addition to the clicker. The last participant benefited from an intervention package that included a clicker, simplified observations, and a signal for when to record data. As could have been expected, there was variability in the amount of support required to improve data-collection practices for these four randomly selected participants. With direct-care staff possessing variable levels of skill or interest in the process of taking data, we expected this to be the case.

The goal of Experiment 1 was to determine the conditions under which we could establish accurate data-collection procedures so that only relevant stimulus changes emitted by the patients would be recorded. As discussed by Fradenburg et al. (1995), characteristics of the observers, nature of the data-collection system, and characteristics of the data-collection

setting have been shown to affect IOA. No studies have shown what combination of all three characteristics needs to be addressed to improve IOA. The current study showed that characteristics of the observers, as addressed through training, improved IOA in Participant 1 only. An intervention addressing characteristics of the observers and the data-collection setting (clickers) improved IOA for Participant 2 only. Participants 3 and 4 needed all three characteristics to be addressed for IOA to be improved. Interestingly, this was only true when the sessions contained less than 10 instances of a target behavior. Across all participants, 15-min intervals in which more than 10 instances of a behavior occurred resulted in low IOA. In instances associated with high-rate behavior, direct-care staff may have greater competing demands (e.g., safety of self and patient) than during situations associated with lower rate behavior. As many of the intervals were associated with lower rate behavior, we focused on improving IOA within this context for the current study. Future research is surely needed to address accuracy of data collection in moments of high-rate behavior, though.

When considering how these data would apply to a larger unit consisting of direct-care staff with varying levels of training in behavioral data collection, we decided to apply a treatment package that consisted of all treatment components. However, we needed to know that some of these components would not interfere with the establishment of effective data-collection practices, as only Participant 4 contacted all intervention components. Thus, one goal of Experiment 2 was to evaluate whether this treatment package would result in increased data-collection accuracy for four additional participants on the psychiatric unit. Also, each participant included in Experiment 1 had several 15-min intervals in which no target behaviors occurred. This may be particularly relevant for Participants 2 and 3, in which at least half of the 15-min intervals collected in the return to the most effective condition contained no problem behaviors. Although we decided to include these data because we felt that it was equally important that only target behaviors were recorded, these 15-min intervals may have been less difficult to code. A second goal of Experiment 2 was to show that accurate data collection would occur with the treatment package when target behaviors occurred during each 15-min interval. To accomplish this, we omitted 15-min intervals in which no target behaviors occurred.

Experiment 2

Method

Participants

Direct-Care Staff (Hereafter Referred to as “Participants”) Four additional participants employed by the specialized psychiatric unit for children diagnosed with intellectual and developmental disabilities participated in Experiment 2. All four

participants had a bachelor's degree in a psychology-related field. All four participants had a brief didactic presentation on data collection but otherwise did not have experience collecting behavioral data. All participants consented to be part of the study and were assured that their job performance would not be evaluated based on participation in this study.

Patients The participants' collection of behavioral data was observed as they worked with patients admitted to the psychiatric unit. Please see Table 1 for demographic information regarding these patients.

Advanced Data Coders The same advanced data coders in Experiment 1 collected data for Experiment 2.

Setting, Materials, Response Definitions, and Observation System The same setting, materials, response definitions, and observation system used during Experiment 1 were also used during Experiment 2. In addition to intervals being excluded because more than 10 instances of a target behavior occurred, intervals were also excluded if no problem behaviors occurred.

IOA On an average of 41% (range 28.6%–83.3%) of 15-min intervals, the second advanced data coder recorded data with the primary advanced data coder. Average IOA was 88.9% (range 0%–100%; see Fig. 2). For Participant 5, IOA was collected on 33% of intervals and was 100%. For Participant 6, IOA was collected on 28.6% of intervals and was 100%. For Participant 7, IOA was collected on 83.3% of intervals and averaged 73.3% (range 0%–100%). Following the interval with 0% agreement, the advanced data coders discussed the disagreements and then observed another 15-min interval for that same participant. Agreement increased following this discussion. For Participant 8, IOA was collected on 33% of intervals and was 100%.

Experimental Design A multiple-baseline across-participants design was used to evaluate differences in IOA following exposure to the treatment package.

Procedures

Baseline (Condition A) The same baseline procedures used in Experiment 1 were used during Experiment 2.

Clicker, Simplified Observation, and Signal (Condition D) The procedures for this condition were the same as described during Experiment 1.

Results and Discussion

IOA for all participants who agreed to participate in Experiment 2 matched the results of the four participants

enrolled in Experiment 1. Average IOA for these four participants at baseline was 14.77% (range 0%–60%). IOA for Participants 5, 6, and 7 during the baseline condition was 0%. Participant 8's IOA during the baseline condition was 36.9% (range 0%–60%).

Upon introduction of the treatment program, all participants' IOA increased. Participant 5's IOA increased to 100% for two consecutive 15-min intervals. Participant 6's IOA was initially variable before stabilizing at 100% for two consecutive 15-min intervals; average IOA was 68% (range 0%–100%). Participant 7's IOA showed an increasing trend before stabilizing at 100% for two consecutive 15-min intervals ($M = 86.7\%$; range 60%–100%). Participant 8's IOA increased to 100% immediately after treatment was introduced. No intervals in which high-rate behavior occurred were selected during this study, despite the advanced data coders continuing to randomly select intervals.

All participants showed initially moderate to low levels of IOA but increased their accuracy upon introduction of the entire treatment package consisting of the clicker, simplified observation, and signal for when data-collection intervals ended. The goal of this experiment was to show that data-collection accuracy could be increased following the introduction of this package. However, given the data collected during Experiment 1, it is likely that a different aspect of the package resulted in improved accuracy for each participant.

General Discussion

The current two-experiment study evaluated the potential interventions that could be introduced to participants working on a psychiatric unit with children diagnosed with intellectual and developmental disabilities to increase accuracy of data collection. During Experiment 1, we found that four participants displayed four different patterns of data. One participant responded well to baseline data-collection measures. The introduction of a clicker resulted in increased data collection for another participant. The other two required a combination of a clicker plus simplified observation (Participant 3) and a combination of those two components and a signal for when an interval ended (Participant 4). We then evaluated whether four additional participants would respond positively to a treatment package consisting of all intervention components (Experiment 2). We found that increased IOA was noted for all four participants.

The American Academy of Child and Adolescent Psychiatry published a position statement that encourages providers working with children diagnosed with autism spectrum disorders to use systematic data collection and behavioral principles to modify problem behaviors (Volkmar et al., 2014). Although recognition is being given for the importance of systematic data collection, position statements like these offer little in the way of guidance for how to establish data-

Fig. 2 Total IOA for Participants 5–8 during Experiment 2

collection systems within programs that historically have not collected data on child behavior. As data-collection practices are being introduced to these organizations, attention needs to be directed to strategies to address interfering factors to accurate data collection (Fradenburg et al., 1995).

In the current study, we systematically evaluated three environmental conditions hypothesized to influence accuracy of data collection (Fradenburg et al., 1995). Specifically, we evaluated characteristics of the observers (training), nature of the data-collection system (simplified observations), and characteristics of the setting (clickers and signals). Results of Experiment 1 were highly individualized. One participant responded to a didactic training program only. The other three participants needed some form of environmental modification to improve performance. It could be that the prompts (i.e., recording data every 15 min) for data collection were not discriminable during baseline (Vollmer et al., 2008). Thus, clicker, simplified observation, and signal may have functioned as more effective prompting strategies to improve performance. Interestingly, this was only the case for intervals in which less than 10 target behaviors occurred. Intervals in which high-rate behavior occurred were always associated with lower IOA during Experiment 1. Future research needs to address this concern. It could be that a more accurate data-collection method during these types of intervals would be to

record an “episode” of target behavior as opposed to recording an estimate of target behavior occurrence.

The current study is not without its limitations. First, given the structure of the specialized unit, the staff changed the patients they worked with daily. Thus, they were observed collecting data with multiple patients. Some patients’ data were likely easier to collect than others. During Experiment 1, data during intervals in which more than 10 instances of behavior occurred were associated with lower IOA for all participants. Future research should investigate ways to improve the accuracy of data collection for staff that simultaneously need to collect data while providing clinical care for patients engaging in high-rate behavior. Second, and as discussed earlier, several of the 15-min intervals evaluated during Experiment 1 contained no target behaviors. Data collection may have been easier for these intervals compared to intervals in which problem behaviors occurred. We addressed this limitation in Experiment 2 but believe that future research should investigate this further. Relatedly, a third limitation of the current study is related to the conditions that omitted certain dependent variables. For example, whenever simplified observation was used, only the most frequently occurring problem behavior was recorded. Although the advanced data coders collected data on the other dependent variables during these studies, this practice may disrupt monitoring of less frequently occurring, though still important, problem behaviors. Future

research should evaluate strategies to increase the number of behaviors observers can accurately track to avoid ignoring these behaviors. A fourth limitation is the relatively small sample of 15-min intervals observed for the participants. This could call into question how robust the current data are. Thus, future research should replicate the current study and provide extended maintenance data to establish the value of these procedures for establishing lasting change.

Implications for Practice

- Discussion of the importance of behavioral data collection for practitioners.
- Description of environmental barriers to accurate data collection.
- Demonstration of a clinical evaluation to evaluate environmental modifications to improve accurate data-collection practices.
- Recommendations for practitioners and researchers interested in establishing accurate data-collection practices.

Acknowledgements The authors would like to thank Lyndsay Gaffey for her support of this project.

Compliance with Ethical Standards

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

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

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

References

- Asmus, J. M., Vollmer, T. R., & Borrero, J. C. (2002). Functional behavior assessment: A school based model. *Education and Treatment of Children*, 25, 67–90.

- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1, 91–97. <https://doi.org/10.1901/jaba.1968.1-91>.
- Blough, D., Pulliam, B., Page, T. J., Sweeney, V., Dougherty, R., & Grim, M. (2006). Improving evaluation of psychotropic medication for adults with developmental disabilities living in community settings. *Behavioral Interventions*, 21, 73–83. <https://doi.org/10.1002/bin.203>.
- Fradenburg, L. A., Harrison, R. J., & Baer, D. M. (1995). The effect of some environmental factors on interobserver agreement. *Research in Developmental Disabilities*, 16, 425–437. [https://doi.org/10.1016/0891-4222\(95\)00028-3](https://doi.org/10.1016/0891-4222(95)00028-3).
- Mash, E. J., & McElwee, J. D. (1974). Situational effects on observer accuracy: Behavioral predictability, prior experience, and complexity of coding categories. *Child Development*, 45, 367–377. <https://doi.org/10.2307/1127957>.
- Mowery, J. M., Miltenberger, R. G., & Weil, T. M. (2010). Evaluating the effects of reactivity to supervisor presence on staff response to tactile prompts and self-monitoring in a group home setting. *Behavioral Interventions*, 25, 21–35. <https://doi.org/10.1002/bin.296>.
- Mozingo, D. B., Smith, T., Riordan, M. R., Reiss, M. L., & Bailey, J. S. (2006). Enhancing frequency recording by developmental disabilities treatment staff. *Journal of Applied Behavior Analysis*, 39, 253–256. <https://doi.org/10.1901/jaba.2006.55-05>.
- Sigurdsson, S. O., & Austin, J. (2006). Institutionalization and response maintenance in organizational behavior management. *Journal of Organizational Behavior Management*, 26, 41–77. https://doi.org/10.1300/j075v26n04_03.
- Tiger, J. H., Hanley, G. P., & Bruzek, J. (2008). Functional communication training: A review and practical guide. *Behavior Analysis in Practice*, 1, 16–23. <https://doi.org/10.1007/BF03391716>.
- Volkmar, F., Siegel, M., Woodbury-Smith, M., King, B., McCracken, J., State, M., & AACAP Committee on Quality Issues. (2014). Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53, 237–257. <https://doi.org/10.1016/j.jaac.2013.10.013>.
- Vollmer, T. R., Sloman, K. N., & St. Peter Pipkin, C. (2008). Practical implications of data reliability and treatment integrity monitoring. *Behavior Analysis in Practice*, 1, 4–11. <https://doi.org/10.1007/BF03391722>.
- Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11, 203–214. <https://doi.org/10.1901/jaba.1978.11-203>.
- Wooderson, J. R., Cuskelly, M., & Meyer, K. A. (2014). A systematic review of interventions for improving the work performance of direct support staff. *Research and Practice in Intellectual and Developmental Disabilities*, 1, 160–173. <https://doi.org/10.1080/23297018.2014.941967>.