International Electronic Journal of Elementary Education, December 2016, 9(2), 385-404.

A Preliminary Analysis of a Behavioral Classrooms Needs Assessment

Justin B. LEAF [a]* Ronald LEAF[a] Cynthia McCRAY [b]
Carol LAMKINS [b] Mitchell TAUBMAN [a] John McEACHIN [a]
Joseph H. CIHON [a]

[a] *Autism Partnership Foundation, USA*
[b] *Clark County Unified School District, USA*

Received: September, 2016 / Revised: October, 2016 / Accepted: November, 2016

Abstract

Today many special education classrooms implement procedures based upon the principles of Applied Behavior Analysis (ABA) to establish educationally relevant skills and decrease aberrant behaviors. However, it is difficult for school staff and consultants to evaluate the implementation of various components of ABA and general classroom set up. In the present study we developed the Behavioral Classroom Needs Assessment as a tool to measure the quality of implementation of principles derived from ABA, teaching, and classroom set up in special education classrooms. Experiment 1 evaluated the reliability of two observers using the Behavioral Classroom Needs Assessment during 128 different observations across 68 different special education classrooms. An Intraclass Correlation Coefficient and Cronbach Alpha Analysis were utilized to determine reliability, and the results showed a high f of reliability across the 40 questions of the assessment. Experiment 2 compared the quality of intervention using the Behavioral Classroom Needs Assessment in five classrooms who received behavioral consultation and five classrooms that did not receive behavioral consultation. The results showed an improvement in the scores on the Behavioral Classroom Needs Assessment for those classrooms in which consultation occurred.

Keywords: Applied behavior analysis, Assessment, Autism, School

Introduction

Applied Behavior Analysis (ABA) is the application of behavioral principles to improve the lives of individuals (Baer, Wolf, & Risley, 1968, 1987). ABA based intervention can be implemented by a wide variety of people, including behavior analysts (Shook, Ala'i-Rosales, & Glenn, 2002), parents (Charlop-Christy & Carpenter, 2000), teachers (Koegel, Russo, & Rincover, 1977), and paraprofessionals (McCulloch & Noonan, 2013). A variety of procedures are implemented under the umbrella of ABA, including but not limited to: reinforcement (Cooper, Heron, & Heward, 2007), prompting (e.g., Leaf, Sheldon, & Sherman, 2010; Touchettte, 1971), functional behavioral assessment (e.g., Hanley, Iwata, & McCord, 2003), punishment (e.g., Lerman & Vorndran, 2002), time-out (e.g., Donaldson

* Corresponding author: Justin B. Leaf, Autism Partnership Foundation, 200 Marina Drive, Seal Beach, CA 90740, United States. Email: Jblautpar@aol.com

ISSN:1307-9298

www.iejee.com

& Vollmer, 2011), the teaching interaction procedure (TIP; e.g., Leaf et al., 2012), token economies (e.g., Ayllon & Azrin, 1965), video-modeling (e.g., Charlop-Christy, Le, & Freeman, 2000), and discrete trial teaching (DTT; Lovaas, 1987).

To date, procedures based upon the principles of ABA have strong empirical support, demonstrating effectiveness in improving the quality of life for individuals diagnosed with ASD (e.g., Howard, Ladew, Pollack, 2009; Odom, Collet-Klingenberg, Rogers, & Hatton, 2010). With the increasing prevalence of ASD, the number of children with diagnosed with ASD within public schools has risen. Moreover, the U.S. Department of Education (2011) reported that the number of students diagnosed with ASD who have had an IEP rose from approximately 95,000 in 2000 to over 450,000 in 2011. These students can be placed in regular education classrooms, self-contained special education classrooms, autism classrooms, or resource classrooms.

The increased prevalence of ASD has resulted in more classrooms using ABA methodology and more teachers and paraprofessionals receiving training in the principles of ABA (Carr, Howard, & Martin, 2015). It is common that teachers receive ongoing supervision and training on behavior analytic procedures by a variety of professionals, which can include on-site behavior analysts, off-site behavioral consultants, autism consultants, school administrators (e.g., principals, vice-principals, counselors), and district administrators (e.g., superintendents, special education directors, or special service supervisors). It is important that those utilizing behavioral analytic procedures implement them with a high degree of treatment fidelity to ensure meaningful progress, prevent errors by the learner, and the potential of harm to the learner. Thus, a comprehensive assessment that measures how well principles of ABA are implemented would be important and useful in school settings.

A comprehensive assessment could be beneficial for several reasons. First, it could be utilized to identify strengths and weaknesses that a classroom has at any given moment, which could be used for feedback and to inform training. Second, it could be utilized to track the progress of school staff throughout the school year. Tracking progress could help indicate if training can be faded or if a classroom requires more training. Third, it could be used in cases of litigation to determine if teachers and paraprofessionals are providing appropriate intervention based upon the student's IEP. Finally, a comprehensive assessment can help ensure that students diagnosed with ASD and other developmental disabilities receive the highest quality of intervention, which will help to ensure the students achieve the best possible outcomes.

Therefore, the purpose of this study was twofold. Experiment 1 explored the development of a Behavioral Classroom Needs Assessment by the researchers along with professional members of a large western school district to evaluate nine different domains pertaining to implementation of the principles of ABA within classrooms with a high degree of quality. Moreover, we evaluated the reliability of two independent observers when using the Behavioral Classroom Needs Assessment while simultaneously observing classrooms throughout the academic school year. Experiment 2 used the Behavioral Classroom Needs Assessment to compare 5 classrooms receiving consultation to 5 classrooms not receiving consultation.

Experiment 1

Method

Setting. This study was conducted in a school district located in the Western part of the United States. The school district had received consultation for 15 years prior to this study and requested participation within the study. The study was conducted in 69 special

education, autism, and resource classrooms, which included classrooms in preschool, kindergarten, elementary, middle school, and high school. Each classroom was set up based upon each individual teacher's preference.

Development of the Behavioral Classroom Needs Assessment. In order to track ongoing progress of the behavioral consultation, the special education director and assistant special education director requested that the researchers to develop an assessment to measure if improvements were observed in the classrooms in which consultation occurred. The researcher (first author), the director of the district's outside consultation agency (second author), district administrators (third and fourth author), and members of the district's in-house consultant who provided ongoing training to teachers collaborated in the development of the assessment.

To help create the assessment the researcher (first author), consultant, one administrator, and members of the in-house consultant met to discuss domains of ABA, what areas are important within the school district, and specific teacher/paraprofessional skills that would fall within each domain. Nine domains were identified, with specific skills within each domain that constituted a well conducted classroom. The domains consisted of: (a) age appropriateness; (b) curriculum; (c) reinforcement; (d) behavior plans (proactive and reactive); (e) teaching strategies; (f) DTT; (g) shadow support; (h) data; and (i) classroom environment.

After the skills were determined, the first author, created operational definitions of each, which corresponded with a five-point Likert scale (contact first author for a list of the operational definitions). At the start of the 2013-2014 school year the first author conducted a one-day training on the Behavioral Classroom Needs Assessment to the in-house consultants. This training consisted of didactic instruction and hands-on practice using the assessment tool. In-house consultants who scored over 85% reliability across all questions of the assessment concluded the training. Members who scored lower than 85% reliability continued with further observations and feedback until reaching at least 85% reliability with the lead researcher, at which point training was concluded.

Behavioral Classroom Needs Assessment. The Behavioral Classroom Needs Assessment consisted of a total of 40 questions, all of which answers were given on a five-point Likert scale. The five-point Likert scale generally consisted of scoring between very rarely displayed (approximately 0 to 24% of the time), rarely displayed (approximately 25 to 59% of the time), sometimes displayed (approximately 60 to 79% of the time), most of the time displayed (approximately 80 to 94% of the time), or almost always displayed (approximately 95% or above). There was also a not applicable score that raters could indicate if a classroom behavior could not be observed. For example, in the paraprofessional support (see below) domain, a rater might mark a score of not applicable if there was no paraprofessional in the room during the observation. The 40 questions were broken into nine larger domains. See Appendix A for a copy of the Behavioral Classroom Needs Assessment.

Observers. The observations (described below) were conducted by 17 in-house consultants and the first author. Each in-house consultants had been a supervisor from 1 to over 5 years. In-house consultants were responsible for training teachers in the principles of ABA, working directly with students, attending IEP meetings, and holding large district wide trainings.

For the purpose of this study, each in-house consultant was randomly assigned classrooms where s/he was responsible for conducting the Behavioral Classroom Needs Assessment. The classrooms were randomly assigned, but the researcher ensured that the in-house consultant did not observe classrooms where s/he provided consultation. Each in-house

consultant was instructed to observe each classroom at least four times throughout the school year and to allow at least three school weeks between observations. A second in-house consultant or the first author was randomly assigned to observe simultaneously, but independently with the primary observer. The second in-house consultant could also not provide supervision to the classrooms for which he or she was observing.

Observation Periods. An observation period lasted approximately 20 min. Observation periods were conducted at random times during the day and during different instructional periods. An observation could only occur if the lead teacher or a long term substitute teacher was present during the observation and if at least a portion of the observation took place in the classroom. During a portion of some observations the entire class went to another location in the school (e.g., music time, gym time, or library time).

The lead teacher of each classroom was informed at the beginning of the school year that observations would occur throughout the school year. The teachers were informed the observations would not affect their job status or be used for feedback purposes. During each observation the in-house consultant asked the teacher if s/he could observe and instructed the teacher to carry on as normal. If the lead teacher did not give assent, the observers attempted to come back at a later time. For the entire observation, the observers attempted to remain as unobtrusive as possible and deferred answering any questions that were posed. The observers did walk around the room to observe what was being taught, look at decorations, and observe data collection procedures. The observers did not talk to the teacher or paraprofessional during the observation. The observers scored the Behavioral Classroom Needs Assessment during the observation and completed all questions within 5 min of the observation.

Dependent Variables and Measures. The main dependent variable was the reliability between the primary observer and the secondary observer on the 128 classroom observations. Overall inter-rater reliability across all of the questions on the Behavioral Classroom Needs Assessment, of each of the nine domains, and each of the questions was assessed. Two different methods for calculating inter-rater reliability were used (described below).

The first assessment used to assess reliability was the Intraclass Correlation Coefficient. The Intraclass Correlation Coefficient is a descriptive statistic which can be used to measure reliability (Yoder & Symons, 2010). The Intraclass Correlation Coefficient identifies "the proportion of total variance in a reliability sample due to between-person variance in the total score" (Yoder, & Symons, 2010 p. 211). We used the Intraclass Correlation Coefficient alpha, with a two way mixed model, using absolute agreement, and utilizing the single measure. Interpretative ranges of the Intraclass Correlation Coefficient are as follows: $a \geq 0.8$ is almost perfect; $0.7 \leq a < 0.79$ is strong agreement; $0.5 \leq a < 0.69$ is moderate agreement; $0.3 \leq a < .49$ is fair agreement; and $0 \leq a < .0.2$ is poor agreement (Fleiss, 1981; Landis & Koch, 1977; Portney & Watkins, 2000).

Second, we calculated the Cronbach Alpha to determine the internal consistency across the two raters. A Cronbach Alpha was calculated per question, per domain, and across all questions of the Behavioral Classroom Need Assessment. The scores from the Cronbach Alpha are interpreted as follows: $a \geq 0.9$ is excellent; $0.7 \leq a < 0.9$ is good; $0.6 \leq a < 0.7$ is acceptable; $0.5 \leq a < 0.6$ is poor; and $a < .49$ is considered unacceptable, as recommended by Geroge and Maller (2003) and Kline (2000).

Results and Discussion

Reliability per Domain and Overall Reliability

Table 1 provides the results of the reliability per domain and the overall reliability across the entire assessment according to the Interclass Correlation Coefficient and Cronbach Alpha. For the Interclass Correlation Coefficient three domains had almost perfect reliability, five had strong reliability, and one had moderate reliability. Across all domains the Interclass Correlation Coefficient yield almost perfect reliability. For the Cronbach Alpha, on domain had excellent reliability, seven had good reliability, and one had acceptable reliability. Across all domains Cronbach Alpha yielded excellent reliability.

Reliability per Question

We conducted an item analysis for each of the 40 questions, using the two measures, to determine the reliability between the primary observer and a secondary observer (see Table 2). The results from the Interclass Correlation Coefficient indicated that 8 (20%) of the questions showed almost perfect reliability, 17 (42.5%) showed strong reliability; 14 (35%) showed moderate reliability, and 1 (2.5%) showed fair levels of reliability. For the Cronbach Alpha analysis, all questions showed an acceptable or higher amount of reliability, with 10% of the questions having excellent reliability, 77.5% of the questions having good reliability, and 12.5% of the questions having acceptable reliability.

The purpose of Experiment 1 was to develop and examine the reliability of independent observers using the Behavioral Classroom Needs Assessment in classrooms providing education for children diagnosed with ASD. The results suggest that independent observers achieved acceptable or higher levels of reliability on each of the 40 questions, 9 domains, and high levels of reliability on the overall assessment. In this experiment, the Behavioral Classroom Needs Assessment was evaluated across several classrooms and with several different raters and yet there was still a high level of reliability. Not only do the results show that reliability was high across the three methods of calculation, but the results showed co-variance across the different observers; both observers were able to adequately agree whether a classroom was a stronger classroom (i.e., scoring higher on the Behavioral Classroom Needs Assessment).

This level of reliability is impressive, given the number of variables which could impact agreement during an observation. For example, the observers were looking across an entire classroom, as opposed to a single person or a single behavior. Yet the two observers achieved reliable results across the majority of questions. Also, although each question had behavioral markers, there was still some level of subjectivity that occurred when scoring each of the questions. For example, although increasing teachable moments was behaviorally defined, there was some level of subjectivity on what was classified as a teachable moment. It could be that one observer believed there were 10 teachable moments while another observer only witnessed eight teachable moments. Despite the level of subjectivity, the observers still demonstrated a high degree of reliability and yield information that has potentially practical value

Despite the overall high degree of reliability, there are several limitations worth noting. First, the items in the classroom environment domain had less than desirable reliability. One possible reason for lower reliability could be that the items were too global in scope and not adequately focused on specific content. A second limitation is the lack of measures of curriculum examined. The researchers elected not to evaluate curriculum as there are other curriculum assessments available (e.g., Leaf & McEachin, 1999; Sundberg, 2008) and it would be difficult to evaluate curriculum within a 20 min observation period. A third limitation is that this study only evaluated the reliability of the assessment and no

measures were taken on the progress of students in the classrooms, how the assessment can be utilized to provide feedback to staff, or if consultation to classrooms could help improve the quality of teaching and thus improve scores on the Behavioral Classroom Needs Assessment.

Experiment 2

The purpose of Experiment 2 was to evaluate the effects of behavioral consultation for the quality of behavioral intervention with classrooms. Effects of consultation was measured suing The Behavioral Classroom Needs Assessment described in Experiment 1.

Methods

Setting. Experiment 2 occurred in the same school district from Experiment 1. A total of 10 classrooms were observed throughout the experiment. Five of which received behavioral consultation in the 2015-2016 school year and five did not. Which classrooms received consultation was not determined by random assignment, but, rather, by the request for consultative services. A request for consultative service can either be made by the teacher him- or herself or by an administrative member (e.g., principle or vice principle) of the school where that classroom was located.

Dependent Variable and Measurement. The main dependent variable in Experiment 2 was the overall percentage of points earned across all nine domains of the assessment for each classroom across four different periods of time. To determine this score, the authors added the total number of points (from the Likert scale) earned across all questions and all teachers within one of the two groups and divided it by the total number of points across all questions and teachers and multiplied by a 100. The second measure was the overall percentage of points earned within each domain for each classroom across four different periods of time. The score for each domain was done identical to how the researchers determined the score across all questions within the assessment.

The first author was the primary scorer of the Behavioral Classroom Needs Assessment. While the first author was not blind to the purpose of this study, he was blind to which classrooms were receiving consultation throughout the school year. A second observer was present during all observations. S/he worked for the school district but was unaware if the classroom was or was not receiving consultation.

Observations took place across four different time periods. The first observation (T1) took place prior to consultation to any of the classrooms. The second observation (T2) took place three months after consultation began for the designated classrooms. The third observation (T3) took place about three months after the second observation. The fourth and final observation (T4) took place one month after the third observation.

Observation Periods

Observation periods were identical to Experiment 1

Consultation. The school district used a pyramidal approach to consultation in which outside consultants were hired to provide consultation to in-district consultants and on occasion to the teachers directly. The in-district consultants then provided direct consultation to the teachers and classrooms on a regular basis (ranging from weekly to monthly consultation). Each of the classrooms assigned to the consultation condition had one in-district consultant assigned for the entire academic year. The in-district consultant would provide direct consultation on a bi-monthly to monthly basis.

There were a total of 6 consultants contracted throughout the year providing 200 of days of consultation across the academic year. Each outside consultant was assigned anywhere

from 1 to 5 different in-district consultants to supervise throughout the year. The outside consultant met with the in-district consultant on a monthly basis. The outside consultation consisted of didactic instruction to the in-district consultant, didactic instruction occasionally to teachers (e.g., workshops that teachers attended), and hands on training within the classroom. There was a wide variety of procedures covered which included: (a) characteristics of autism; (b) principles of reinforcement; (c) DTT; (d) curriculum objectives; (e) classroom management; (f) behavior intervention plan development; (g) principles of ABA; and (h) effectively communicating with the teachers. It was the responsibility of the in-district consultant to take the concepts taught and apply them to directly to the classrooms. All areas where trained using a progressive model of ABA (Leaf et al., 2016) in which they taught the in-district consultants and teachers to use clinical judgments, make in-the-moment decisions, and be guided by the students as opposed to following a strict protocol.

There were are a total of 15 in-district consultants, five of which were assigned to the five classrooms in this condition. The in-district consultant provided direct training to the classroom teachers. This training consisted of modeling correct instructor behaviors, providing coaching during the regular school day, and providing didactic instruction when the teacher was not working with any of her or his students. The in-district consultant covered the same topics that the outside consultant worked on throughout the school year.

Reliability. The primary scorer used the Behavioral Classroom Needs Assessment during every observation to evaluate the classroom. On 52% of sessions the secondary observer also scored Behavioral Classroom Needs Assessment to assess reliability. Intraclass Correlation Coefficient was used to assess reliability (Yoder & Symons, 2010). We evaluated the reliability across all 40 questions of the assessment. The results showed that 47.5% (19 questions) reliability was almost perfect, 27.5% (11 questions) reliability was strong, 20% (8 questions); 5% (2 questions) reliability was fair, and 0% of the questions reliability was poor. The only two questions with fair reliability were if the shadow teacher provided proper amount of reinforcement and if materials were organized.

Data Analysis. We conducted a one-tailed, independent sample t-test to evaluate whether scores for the group that received consultation were statistically different from the scores for the group that did not receive consultation. We compared the two groups on their overall score and their scores on each of the domains across the four periods of time (T1, T2, T3, and T4).

Results and Discussion

The main dependent variable for this study was comparing the overall increase in the score on Behavioral Classroom Needs Assessment for each classroom. Figure 1 displays these results. Across the x-axis is the four different time period of observation and across the y-axis is the percentage of points earned. The closed circles represents data for the teachers/classrooms who received consultation and open squares represent data for the teachers/classrooms who did not receive consultation.

On T1 the average score across the five classrooms that received consultation was 31.8% (range 22.3 to 42.9% across classrooms) and 24.3% (range 13.2 to 42.3% across classrooms) across the five classrooms that did not receive consultation. The results of the one-tailed t-test showed no significant difference between the two groups ($t(8) = 1.262$, $p = .121$). On T2 the average score across the five classrooms that received consultation was 46.3% (range 20.9 to 74.3% across classrooms) and 31.7% (range 10.1 to 73.6% across classrooms) across the five classrooms that did not receive consultation. The results of the one-tailed t-test showed no significant difference between the two groups ($t(8) = 1.003$, p

=.173). On T3 the score across the five classrooms that received consultation was 57.8% (range 36.1 to 82.3% across classrooms) 37.2% (range 12.8 to 73% across classrooms) across the five classrooms that did not receive consultation. The results of the one-tailed t-test showed no significant difference between the two groups ($t(8)$ =1.4.16, p =.173). During T4 the average score across the five classrooms that received consultation was 65.7% (range 22.3 to 42.9% across classrooms) 36.9% (range 17.5 to 63.4% across classrooms) across the five classrooms that did not receive consultation. The results of the one-tailed t-test showed a significant difference between the two groups ($t(8)$ =2.3, p =0.0275).

The results of the main dependent variable could be interpreted as follows. First, across the five classrooms who received consultation an increasing trend can be observed across the four observation periods. Moreover, from T1 to T4 there was an increase in 30 percentage points. Second, there was a slight increasing trend across the first three observation periods followed by a decrease on observation four with the five classrooms that did not receive consultation. However, the increase was only 12 points compared to the 30 percentage points gained by the classrooms that received intervention. Third, individual differences across classrooms for both groups of classrooms were observed. One classroom that did not receive consultation did show a large gain in percentages points; however, the average across the five classrooms was much less than the classrooms that did receive intervention. Finally, by the end of the academic year there was a significant difference for those classrooms who received consultation to those classrooms who did not receive consultation. Thus, the overall the results showed that consultation could increase the quality of ABA intervention implemented within special education classrooms based on the Behavioral Classroom Needs Assessment.

The second dependent variable was an analysis of scores within each domain. Table 3 provides the score and standard deviation for each domain across the four observation periods. Additionally, we have reported when there was a significant difference between the two groups across each of the domains and across all four periods of time. The results are similar to the overall results in that: (a) classrooms that received consultation showed a higher increase across the domains as observation periods continued; (b) classrooms that did not receive consultation showed some increases across the domains; (c) individual differences could be seen across classrooms; (d) there were significant difference found during certain periods of time and on certain domains; and (e) when there was no statistical significance found there was still clinically significant levels found. Finally, there was no real increase in the domain of data collection within or across the two groups.

The purpose of Experiment 2 was to examine the effects of behavioral consultation using a progressive model of ABA (Leaf et al., 2016) on the fidelity of implementing procedures based upon ABA. The results of Experiment 2 showed that consultation on the principles of ABA resulted in the greater quality intervention than classrooms who did not receive consultation based on the Behavioral Classroom Needs Assessment.

Despite the increase observed for classrooms that received consultation there are some limitations that future research may address. First, a pure random assignment for classrooms was not used, and classrooms that were part of the consultation condition were done so at the request of the teacher or school administration. Thus, the results could potentially be biased as teachers who may select consultation are more amenable to training. Although, this is a methodological flaw, this was done due to ethical and logistical reasons that the school district could not decide randomly which classrooms could and could not receive consultation. Second, this study included a small number of classrooms therefore it is not known what the results would be on a larger scale. Third, there were

individual differences observed across classrooms and future researchers should further evaluate these individual differences. Finally, although improvements were observed across the domains the overall averages were typically under 80%. Thus, improvements were made but not at the level where the teachers would not require ongoing supervision and consultation.

General Discussion

The purpose of this study was twofold. First, to develop and examine the reliability of the Behavioral Classroom Needs Assessment (Experiment 1). Second, to examine the effects of behavioral consultation using the Behavioral Classroom Needs Assessment on the implementation of the principles of ABA. The results from Experiment 1 and 2 suggest that independent observers acceptable or higher levels of reliability on each of the 40 questions, 9 domains, and the overall assessment were obtainable. The results from Experiment 2 suggest that behavioral consultation could lead to better implementation of ABA based procedures within a special education classroom based on the Behavioral Classroom Needs Assessment.

This study contributes to the literature base in at least two ways. First, prior to this study, the majority of assessments for the application of ABA have either been described as part of curriculum books (e.g., Leaf & McEachin, 1999) or have been auxiliary measures as part of larger research studies (e.g., Virsmara et al., 2009). Thus, there have been few, if any, studies that have evaluated an entire classroom through the use of a comprehensive assessment. Second, there have been no studies to date that have evaluated the reliability of two observers on a comprehensive behavioral classroom assessment.

This study has several practical implications for clinicians, teachers, classrooms, and school districts. For children diagnosed with ASD to make meaningful progress they need to receive quality ABA services (e.g., Leaf et al., 2016; Lovaas, 1987). This assessment could be used by consultants and/or school administrators to determine what areas are provided with a high degree of quality and what areas can be improved upon. The effective use of this assessment could result in consultants or administrators providing objective feedback, and used as a tool to help improve the overall classroom. Additionally, parents could use this tool to determine if their child is receiving high quality services. As a result, the tool can help improve the lives for individuals diagnosed with ASD.

Despite the overall high degree of reliability and effectiveness of school consultation future research is still warranted. First, although the results across the two studies were positive, there was no measure if a high score on the Behavioral Classroom Needs Assessment actually resulted in increased student performance or decreased aberrant behavior. Thus, future researchers should conduct an analysis to see if improvement in scores on the Behavioral Classroom Needs Assessment are correlated with better student outcomes. Second, although the preliminary findings show that behavioral consultation was effective this study used a small number of participants. Thus, future researchers may wish to evaluate the Behavioral Classroom Needs Assessment in a large scale random control trial. Finally, evaluators of the classrooms were part of the group that helped design the assessment; therefore, it is unknown how reliable the scale would be for observers who were not part of the assessment's creation. Additionally, all observers were well trained in the principles of ABA. Therefore, it is not known how reliable an observer who has less expertise would be in conducting this assessment.

Despite these limitations, this study has important implications for clinical practice. As more children are receiving diagnoses of ASD and as more classrooms are orienting towards an ABA methodology, it is important that researchers develop instruments that can be used to assess the quality of intervention being implemented inside a classroom

environment. It is also important that behavior analysts evaluate their consultative services within school districts/classrooms to ensure that consultation leads to quality implementation of procedures based upon ABA. This study was a first step toward developing an evaluation and analyzing that evaluation within a large school district. The preliminary results showed a high degree of reliability and that consultation could lead to better implementation of a ABA using a progressive model.

Conflict of Interest

The second, fifth, and sixth authors own an agency that provides consultative services to the school district where this study was conducted. The first and last authors are employed by this agency in which part of their job responsibilities include consulting for school districts.

• • •

References

Ayllon, T., & Azrin, N. (1965). *The token economy: A motivational system for therapy and rehabilitation.* New York, New York: Appleton-Century-Crofts.

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.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis, . *Journal of Applied Behavior Analysis* , *20*(4), 313–327.

Carr, J. E., Howard, J. S., & Martin, N. T. (2015, May). *An updated on the behavior analyst certification board. Panel Discussion presented at the Association for Behavior Analysis International 41st Annual Convention.* San Antonio, Texas.

Charlop-Christy, M. H., & Carpenter, M. H. (2000). Modified incidental teaching sessions: A procedure for parents to increase spontaneous speech in their children with autism. *Journal of Positive Behavior Interventions, 2.* 98-112.

Charlop-Christy, M. H., & Le, L., & Freeman, K. A. (2000). A comparison of video modeling with in vivo modeling for teaching children with autism. *Journal of Autism and Developmental Disorders, 30,* 537-552.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied Behavior Analysis* (2nd Ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Donaldson, J. M., & Vollmer, T. R. (2011). An evaluation and comparison of time-out procedures with and without release contingencies. *Journal of Applied Behavior Analysis, 44,* 693-705.

Fleiss, J. (1981). *Statistical methods for rates and proportions.* New York, NY, USA: Wiley VCH. Garfield,E.

Hanley, G. P., Iwata, B. A., & McCord, B. E. (2003). Functional analysis of problem behavior: A review. *Journal of Applied Behavior Analysis, 36,* 147-185.

Howard, H. A., Ladew, P., & Pollack, E. G. (2009). *The national autism center's national standard project: Findings and conclusions.* Randolph, MA: National Autism Center.

Kline, P. (2000). *The handbook of psychological testing (2nd ed.).* London: Routledge

Koegel, R. L., Russo, D. C., & Rincover, A. (1977). Assessing and training teachers in the generalized use of behavior modification with autistic children. *Journal of Applied Behavior Analysis, 10,* 197-205.

Koegel, L. K., Carter, C. M., & Koegel, R. L. (2003). Teaching children with autism self-initiations as a pivotal response. *Topics in Language Disorders, 23,* 134-145.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159-174.

Leaf, R.B., & McEachin, J.J. (1999). *A Work in Progress: Behavior management strategies and a curriculum for intensive behavioral treatment of autism.* New York, NY: Different Roads to Learning.

Leaf, J. B., Sheldon, J. B., & Sherman, J. A. (2010). Comparison of simultaneous prompting and no-no prompting in two-choice discrimination learning with children with autism. *Journal of Applied Behavior Analysis, 43,* 215-228.

Leaf, J. B., Oppenheim-Leaf, M. L., Call, N. A., Sheldon, J. B., Sherman, J. A., Taubman, M., McEachin, J., Dayharsh, J. & Leaf, R. (2012). Comparing the teaching interaction procedure to social stories for people with autism. *Journal of Applied Behavior Analysis, 45,* 281-298.

Leaf, J.B., Dotson, W. H., Oppenheim-Leaf, M. L., Sherman, J.A. Sherman, J. A. (2012). A programmatic description of a social skills group for young children with autism. *Topics in Early Childhood Special Education, 32,* 111-121.

Leaf, J. B., Leaf, R., Taubman, M., McEachin, J., & Delmolino, L. (2014). Comparison of flexible prompt fading to error correction for children with autism spectrum disorder. *Journal of Developmental and Physical Disabilities, 26,* 203-224.

Leaf, J. B., Leaf, R., McEachin, J., Taubman, M., Ala'i-Rosales, S., Ross, R. K., Smith, T, & Weiss, M. J. (2016). Applied Behavior Analysis is a Science and, Therefore, Progressive. *Journal of Autism and Developmental Disorders, 46,* 720-731.

Lerman, D. C., & Vorndran, C. M. (2002). On the status of knowledge for using punishment implications for treating behavior disorders. *Journal of Applied Behavior Analysis, 35,* 431-464.

Lovaas, I. O. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55,* 3-9.

McCulloch, E. B., & Noonan, M. J. (2013). Impact of online training videos on the implementation of mand training by three elementary school paraproffesionals*. Education and Training in Autism and Developmental Disabilities, 48,* 132-141.

Odom, S. L., Collet-Klingenberg, L., Rogers, S. J., & Hatton, D. D. (2010). Evidence-based practices in interventions for children and youth with autism spectrum disorders. *Preventing School Failure, 54,* 275-282.

Portney, L. G., & Watkins, M. P. (2000).*Foundations of clinical research applications to practice.* Prentice Hall Inc: New Jersey.

Rogers, S. J., & Dawson. (2009). *Play and engagement in early autism: The Early Start Model. Volume 1: The Treatment.* New York: Guilford Press.

Shook, G. L., Alai'i-Rosales, S., & Glenn, S. S. (2002). Training and certifying behavior analysts. *Behavior Modification, 26,* 27-48.

Sundberg, M. L. (2008). *Verbal behavioral milestones assessment and placement program: the VB-MAPP.* Concord: AVB Press.

Taubman, M. T., Leaf, R., McEachin, J., Papovich, S., & Leaf, J. B. (2013). A comparison of data collection techniques used with discrete trial teaching. *Research in Autism Spectrum Disorders, 7,* 1026-1034.

U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS) (2011). Children with disabilities receiving special education under part b of the Individuals with Disabilities Education on Act, 1999-2011.

Vismara, L. A., Colombi, C., & Rogers, S. J. (2009). Can one hour per week of therapy lead to lasting changes in young children with autism? *Autism, 13,* 93-115.

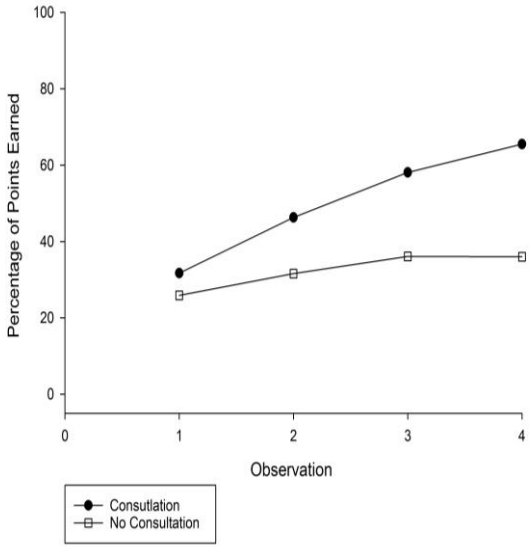

Figure 1. *Overall Score on the Behavioral Classroom Needs Assessment Across the Two Classrooms*

Table 1. *Overall Reliability and Reliability Per Domain*

Domain	Interclass Correlation Coefficient	Cronbach Alpha
Age Appropriateness	0.757 (Strong)	0.862 (Good)
Curriculum	0.713 (Strong)	0.832 (Good)
Reinforcement	0.809 (Almost Perfect)	0.895 (Good)
Behavioral Plan	0.729 (Strong)	0.843 (Good)
Teaching Strategies	0.802 (Almost Perfect)	0.890 (Good)
Discrete Trial Teaching	0.749 (Strong)	0.856 (Good)
Shadow Support	0.769 (Strong)	0.869 (Good)
Data	0.845 (Almost Perfect)	0.916 (Excellent)
Classroom Environment	0.528 (Moderate)	0.691 (Acceptable)
Overall	0.838 (Almost Perfect)	0.912 (Excellent)

Table 2. *Reliability Per Question*

Question Number	Domain	Question	Interclass Correlation Coefficient	Cronbach Alpha
1	Age Appropriate	Reinforcers	0.893 (Almost Perfect)	0.943 (Excellent)
2	Age Appropriate	Language	0.573 (Moderate)	0.730 (Good)
3	Age Appropriate	Curriculum	0.644 (Moderate)	0.782 (Good)
4	Age Appropriate	Interactions	0.658 (Moderate)	0.793 (Good)
5	Age Appropriate	Decorations	0.737 (Strong)	0.848 (Good)
6	Age Appropriate	Instructional Materials	0.745 (Strong)	0.853 (Good)
7	Curriculum	Individualization	0.713 (Strong)	0.832 (Good)
8	Reinforcement	Contingent	0.790 (Strong)	0.884 (Good)
9	Reinforcement	Social Label	0.890 (Almost Perfect)	0.895 (Good)
10	Reinforcement	Schedule	0.763 (Strong)	0.864 (Good)
11	Reinforcement	Differential	0.453 (Fair)	0.643 (Acceptable)
12	Reinforcement	Variety	0.770 (Strong)	0.870 (Good)
13	Reinforcement	Condition	0.797 (Strong)	0.886 (Good)
14	Behavior Plan	SR+ Absence	0.743 (Strong)	0.852 (Good)
15	Behavior Plan	Inadvertent Reinforcement	0.685 (Moderate)	0.814 (Good)
16	Behavior Plan	Least Amount of Attention	0.620 (Moderate)	0.764 (Good)
17	Behavior Plan	Teaching Replacement Behavior	0.548 (Moderate)	0.724 (Good)
18	Behavior Plan	Correct Function	0.722 (Strong)	0.839 (Good)
19	Behavior Plan	Across Settings	0.833% (Almost Perfect)	0.908 (Excellent)
20	Teaching Strategies	ABA Only	0.936 (Almost Prefect)	0.967 (Excellent)
21	Teaching Strategies	Maximize Teaching Trials	0.730 (Strong)	0.845 (Good)
22	Teaching Strategies	Maximize Teaching Moments	0.681 (Moderate)	0.810 (Good)
23	Teaching Strategies	Teacher Style	0.611 (Moderate)	0.763 (Good)

Table 2 (Cont.). *Reliability Questions*

Question Number	Domain	Question	Interclass Correlation Coefficient	Cronbach Alpha
24	Teaching Strategies	Paraprofessional Style	0.636 (Moderate)	0.776 (Good)
25	Discrete Trial Teaching	Close Out Trials	0.711 (Strong)	0.832 (Good)
26	Discrete Trial Teaching	Discrete	0.762 (Strong)	0.864 (Good)
27	Discrete Trial Teaching	Systematic Instructions	0.704 (Strong)	0.830 (Good)
28	Discrete Trial Teaching	Prompting	0.702 (Strong)	0.824 (Good)
29	Discrete Trial Teaching	Pacing	0.740 (Strong)	0.850 (Good)
30	Paraprofessional	Behavior Management	0.786 (Strong)	0.879 (Good)
31	Paraprofessional	Reinforcement	0.818 (Almost Perfect)	0.899 (Good)
32	Paraprofessional	Re-direction to the teacher	0.690 (Moderate)	0.815 (Good)
33	Paraprofessional	Prompting	0.538 (Moderate)	0.697 (Acceptable)
34	Paraprofessional	Rubber band	0.702 (Strong)	0.827 (Good)
35	Data	Performance	0.815 (Almost Perfect)	0.897 (Good)
36	Data	Behavior	0.883 (Almost Perfect)	0.938 (Excellent)
37	Data	Meaningful	0.833 (Almost Perfect)	0.840 (Good)
38	Classroom Environment	Materials	0.511 (Moderate)	0.700 (Acceptable)
39	Classroom Environment	Room Set Up Functionally	0.509 (Moderate)	0.693 (Acceptable)
40	Classroom Environment	Decor	0.561 (Moderate)	0.723 (Acceptable)

Table 3. *Percentage of Points Earned Across Domains and Periods of Time*

Domain	Group	T1	T2	T3	T4
Age Appropriate	Consultation	48.3%	74.9%	86.6%	89.6%
		(14.3)	(26.5)	(12.9)	(13.8)
	No Consultation	46.4%	55.8%	69.1%	74.1%
		(16.2)	(31.8)	(31.8)	(28)
Curriculum	Consultation	35%	65%	95%*	87.5%*
		(33.5)	(37.9)	(11.1)	(14.3)
	No Consultation	30%	44.8%	45%	35%
		(41)	(40.8)	(41.1)	(22.6)
Reinforcement	Consultation	26.7%	48.3%	50.7%	71.8%*
		(14)	(20.1)	(32.6)	(18.4)
	No Consultation	14.5%	32.6%	30.3%	33.3%
		(6.7)	(31.3)	(33.7)	(26.8)
Behavior Plan	Consultation	27.4%	25.8%	37.6%*	50%
		(19)	(30.9)	(26.9)	(39.2)
	No Consultation	13.2%	19.5%	13%	13.2%
		(10.7)	(38.1)	(18.9)	(26.2)
Teaching Strategies	Consultation	39.5%	51.9%	62%	66.3%
		(15.6)	(27.8)	(29.7)	(25.6)
	No Consultation	39%	39.5%	45%	41.9%
		(12)	(25.2)	(32)	(24.5)
DTT	Consultation	24%	33%	61.2%	62.5%*
		(15.1)	(27.7)	(34.9)	(25.9)
	No Consultation	16%	23%	27%	21%
		(17.4)	(38)	(39.8)	(22.1)
Shadow Support	Consultation	16.7%	35.3%	33.8%	61.6%*
		(7.6)	(30.7)	(25.0)	(27.1%)
	No Consultation	15.7%	15.4%	27.5%	19.4%
		(12.7)	(21.2)	(21.7)	(26.7)
Data	Consultation	0%	0%	0%	10.4%
					(20.8)
	No Consultation	3.3%	0%	2.1%	6.6%
		(7.5)		(4.1)	(10.8)
Classroom Environment	Consultation	59.9%	66.6%	81.6%	83.3%
		(6.9)	(11.7)	(13.6)	(0)
	No Consultation	55%	56.6%	64.9%	69.9%
		(17.4)	(25.9)	(22.4)	(23.8)
Overall	Consultation	31.8%	46.3%	57.8%	65.7%*
		(7.9)	(20.5)	(21.5)	(19.3)
	No Consultation	24.3%	31.7%	37.2%	36.9%
		(10.8)	(25)	(24.2)	(18.3)

* Significant at the 0.05 probability level.

Appendix A

Behavioral Classroom Needs Assessment

Date of Observation	
Classroom Observed	
Classroom Ages and Grades	
Lead Teacher	
Number of Students	
Number of Teachers	
Observation Conducted By	
Time of Observation	

Age Appropriateness

		It is Never Age Appropriate? 0 to 24%	Is it Rarely Age Appropriate? 25 to 59%	Is it Sometimes Age Appropriate? 60 to 79%	Is it Most of the Time Age Appropriate? 80 to 94%	Is it Always Age Appropriate? 95 to 100%	N/A
1	Reinforcers	0	1	2	3	4	N/A
2	Language	0	1	2	3	4	N/A
3	Interactions	0	1	2	3	4	N/A
4	Curriculum	0	1	2	3	4	N/A
5	Decorations	0	1	2	3	4	N/A
6	Instructional Materials	0	1	2	3	4	N/A

Curriculum

1	Individualized	Not Individualized 0	For Few Students 1	For Some Students 2	For Most Students 3	For All Students 4	N/A

Reinforcement

		Never 0 to 24%	Rarely 25 to 59%	Somewhat 60 to 79%	Most of the Time 80 to 94%	Always 95 to 100%	N/A
1	Contingent	0	1	2	3	4	N/A
2	Paired with Social and Label	0	1	2	3	4	N/A
3	Appropriate Schedule	0	1	2	3	4	N/A
4	Differential	0	1	2	3	4	N/A
5	Variety of SR+ Used	0	1	2	3	4	N/A
6	Good Condition	0	1	2	3	4	N/A

Behavior Plan (Proactive and Reactive)

		Never 0 to 24%	Rarely 25 to 59%	Somewhat 60 to 79%	Most of the Time 80 to 94%	Always 95 to 100%	N/A
1	Teacher SR+ for absence of aberrant behavior	0	1	2	3	4	N/A
2	Teacher Never Inadvertently SR+ aberrant behavior	0	1	2	3	4	N/A
3	Least Amount of Attention During aberrant behavior	0	1	2	3	4	N/A
4	Teaching of Replacement Behaviors	0	1	2	3	4	N/A
5	Targeting Correct Function	0	1	2	3	4	N/A
6	Across Settings Observed	0	1	2	3	4	N/A

Teaching Strategies

		Never 0 to 24%	Rarely 25 to 59%	Somewhat 60 to 79%	Most of the Time 80 to 94%	Always 95 to 100%	N/A
1	Only ABA Strategies	No (Mark as a 0)			Yes (Mark as a 4)		
2	Maximize # of Teaching Trials	0	1	2	3	4	N/A
3	Captures all Teachable Moments	0	1	2	3	4	N/A
4	Engaging Style (Lead Teacher)	0	1	2	3	4	N/A
5	Engaging Style (Shadow Teachers)	0	1	2	3	4	N/A

Discrete Trial Teaching

		Never 0 to 24%	Rarely 25 to 59%	Somewhat 60 to 79%	Most of the Time 80 to 94%	Always 95 to 100%	N/A
1	SR and SP always provided	0	1	2	3	4	N/A
2	Each trial is discrete	0	1	2	3	4	N/A
3	Instructions are systematic	0	1	2	3	4	N/A
4	Effectively Providing and Fading Prompts	0	1	2	3	4	N/A
5	Appropriate Pace	0	1	2	3	4	N/A

Shadow Support

		Never 0 to 24%	Rarely 25 to 59%	Somewhat 60 to 79%	Most of the Time 80 to 94%	Always 95 to 100%	N/A
1	Provide proper amount of behavior support	0	1	2	3	4	N/A
2	Provide proper amount of reinforcement	0	1	2	3	4	N/A
3	Redirects to the lead teacher when applicable	0	1	2	3	4	N/A
4	Effectively Providing and Fading Prompts	0	1	2	3	4	N/A
5	Overall ability to be a rubber band shadow	0	1	2	3	4	N/A

Data

		Not Taken 0 to 24%	Across a Few Children 25 to 59%	Across Some Children 60 to 79%	Across Multiple Children 80 to 94%	Across All Children 95 to 100%	N/A
1	Data Taken on Performance	0	1	2	3	4	N/A
2	Data Taken on Behaviors	0	1	2	3	4	N/A
3	Meaningful Data	0	1	2	3	4	N/A

Classroom Environment

		Never 0 to 24%	Rarely 25 to 59%	Somewhat 60 to 79%	Most of the Time 80 to 94%	Always 95 to 100%	N/A
1	Materials Organized	None of the Materials are Organized	Few of the Materials are Organized	Some of the Materials Are Organized	Most of the Materials Are Organized	All Materials Are Organized	N/A
2	Room Set Up Functionally	None of the Room Set Up Functionally	Little of the Room Set Up Functionally	Some of the Room Set Up Functionally	Most of the Room Set Up Functionally	Entire Room Set Up Functionally	N/A
3	Overall Décor	Unpleasant Décor	Below Average Decor	Average Décor	Good Décor	Excellent Décor	N/A

This page is intentionally left blank