Spring 4-23-2019

Using Video-Based Training to Teach Students the Conservative Dual-Criteria Method

Chandler Pelfrey
cpelfrey@rollins.edu

Using Video-Based Training to Teach Students the Conservative Dual-Criteria Method

A Thesis
By
Chandler Pelfrey, BS, RBT

Submitted to the Faculty of the Department of Health Professions
at Rollins College in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF ARTS IN APPLIED BEHAVIOR ANALYSIS AND CLINICAL SCIENCE

April, 2019
Winter Park, FL

Acknowledgements

I would like to thank my family, friends, cohort, supervisors, and advisors for their unflagging support. In particular, I would like to thank Dr. Sarah Freeman and Dr. Kara Wunderlich for their patience and encouragement as well as Kristal Huffman BCaBA and Melanie Agrachov BCaBA for their expertise and time. I would like to thank Dr. Jesse Dallery for his help in assessing autocorrelation. Finally, I would like to thank Dr. Christopher M. Swoboda, Dr. Thomas R. Kratochwill, and Dr. Joel R. Levin for allowing me to use their conservative dual-criteria Excel sheet in the current studies.

Table of Contents

Abstract

Practicing behavior analysts and registered behavior technicians (RBTs) are required to base their practices on scientific evidence. Analyzing single-subject data is among the most important behavior-analytic practices because it guides clinical decision-making. Although prior research has shown the conservative dual-criteria (CDC) method is accurate (Fisher, Kelley, & Lomas, 2003) and can be used with real graphs like those used in behavior-analytic practice (Lanovaz, Huxley, & Dufour, 2017; Wolfe, Seaman, Drasgow, & Sherlock, 2018), most empirically supported training procedures involve in-person training. Because in-person training in behavior-analytic practice can be expensive, remotely administered training packages might be more viable. Using a multiple baseline design, we evaluated the efficacy of a remotely administered, video-based training package to teach graduate students to implement the CDC method across two studies. The video-based training included written instructions and a video model. Subjects received packets of AB graphs and interpreted them by answering "yes" or "no" to the question, "Does the graph show a treatment effect?" We measured the accuracy of each subject's interpretations as well as his or her correspondence with the CDC method before and after the video-based training. Limitations of the CDC method and future research directions are discussed.

Keywords: conservative dual-criteria method, Type I error, visual inspection

Introduction

Behavior analysts primarily use unassisted visual inspection to analyze single-subject data. Although individual applications of visual inspection likely vary, behavior analysts are trained to implement some core practices when conducting visual inspection. The standard principles textbook for applied behavior analysis (ABA) outlines three core practices that comprise visual inspection: analyzing data within phases, comparing data between or across phases, and applying steady state strategy (Cooper, Heron, & Heward, 2007).

For the first two practices—analyzing data within and across phases—behavior analysts attend to the level, trend, and variability of behavior. Level refers to the amount of behavior (e.g., 12 responses per hour), trend refers to whether the level of behavior is increasing, decreasing, or staying the same over time, and variability refers to the degree of difference between two or more data points (i.e., the amount of stability in a data set). The last core practice, steady state strategy, involves applying a set of logical assumptions to the patterns observed within and across phases. Briefly, the assumptions are that if behavior continued under baseline conditions, no significant changes would occur and if treatment is effective, then behavior will change when treatment is implemented. Behavior analysts, especially when conducting research, use experimental designs that expose behavior to both baseline and treatment multiple times to replicate both the effect of treatment across conditions and verify behavior patterns within conditions. Behavior analysts might consider additional variables when deciding whether their treatments were effective. For example, one behavior analyst might rate a treatment that reduced a client’s self-injury by 50% as effective but another might require a reduction of 80% or more due to the severity of the behavior. Clients’ opinions about the acceptability of the treatment procedures and importance of the observed behavior change might also inform behavior analysts' interpretations.

Because different visual inspectors might implement these core practices differently, the reliability of visual inspection can be low. Although research evaluating the reliability of visual inspection has shown mixed results (Matyas & Greenwood, 1990; Kahng et al., 2010), some have argued reliable statistical methods should be used to analyze single-subject data, especially when intervention effects are small (Gentile, Roden, & Klein, 1972) and when data are variable or autocorrelated (Matyas & Greenwood, 1990). Others have argued statistical methods might place harmful, hypothetical assumptions on the data (Sidman, 1960), replace behavior analysts' focus on social significance with a focus on statistical significance (Michael, 1974), or be too difficult to translate to clinical practice (Barlow, Nock, & Hersen, 2009).

According to the Behavior Analyst Certification Board's Ethical and Compliance Code, behavior analysts and behavior technicians are required to base their practices in scientific evidence (BACB, 2014). Analyzing single-subject data is one of the most important practices because it informs analysts' clinical decisions. A few empirically tested data analysis methods for single-subject data are traditional visual inspection (Kahng et al., 2010; Matyas & Greenwood, 1990; Stewart, Carr, Brandt, & McHenry, 2007; Wolfe & Slocum, 2015), the split-middle method (Fisher et al., 2003), statistical modeling (Fisher et al., 2003; Huitema, 2004), and the conservative dual-criteria (CDC) method (Fisher et al., 2003; Stewart et al., 2007; Lanovaz et al., 2017; Wolfe et al., 2018).

Review of Literature

Visual Inspection

Prior research has shown the reliability and accuracy of visual inspection might vary due to training. Matyas and Greenwood (1990) assessed the reliability and accuracy of students' interpretations of 27 simulated AB graphs. The authors varied the programmed effect size, variability, and autocorrelation (i.e., the degree of similarity between adjacent data points) in the data sets to assess their effects on students' interpretations. Matyas and Greenwood found more students made false-positive errors (i.e., incorrectly detecting a treatment effect) when autocorrelation and variability were programmed into the graphs. False-positive error rates ranged from 16% to 84% across these parameters. Although the high false-positive rates observed in this study suggest visual inspection can be unreliable and inaccurate, later studies put this finding in context.

Fisher et al. (2003) argued that because Matyas and Greenwood (1990) only used one graph for each combination of parameters, their sample of graphs did not represent the intended behavior patterns. For example, because Matyas and Greenwood's formula included "random" variability, the actual autocorrelation for a graph with a programmed autocorrelation of 0.3 might be much higher (e.g., 0.55) or lower (e.g., 0.11). These discrepancies between the actual data patterns and programmed data patterns, combined with the small number of graphs used, limit the generality of the findings. Further, Kahng et al. (2010) assessed the reliability of expert visual inspectors' interpretations of 36 simulated ABAB graphs. They found high interrater agreement (mean kappa was .84) suggesting replication via a convincing experimental design or inspectors' level of training (students in the Matyas and Greenwood study versus experts in the Kahng et al. study) might influence the reliability and accuracy of visual inspection.

The few studies that focused on training visual inspection have shown mixed outcomes. Wolfe and Slocum (2015) evaluated the effects of two types of instruction (computer-based and lecture-based) on students' accuracy when using visual inspection to interpret simulated AB graphs. Results showed the accuracy of both the computer-based group and the lecture-based group increased similarly following instruction whereas a control group's accuracy did not. Although these results suggest training improves the accuracy of visual inspection, Stewart et al. (2007) found a brief lecture-based training did not improve the accuracy of students' visual inspection. However, both types of instruction in the Wolfe and Slocum study lasted 105 min compared to Stewart et al.'s lecture which lasted only 12 min. Thus, only in-depth visual inspection training might be effective. More research is needed to assess the accuracy of visual inspection in clinical practice and identify effective visual inspection training procedures.

Split-Middle Technique

The split-middle technique involves estimating a line of regression for each phase of single-subject data to aid visual inspection. Some advantages of this method are it can be conducted quickly without a computer and might improve inspectors' analysis of trend (Barlow et al., 2009; Cooper et al., 2007). However, a comparison of multiple data analysis techniques across 30,000 simulated AB graphs, conducted by Fisher et al. (2003), showed the split-middle technique, combined with a binomial test, produced high false-positive rates across multiple phase lengths and autocorrelation parameters (>20% in all cases). This suggests the split-middle technique might only be appropriate when used in combination with steady state strategy, rather than thresholds for statistical significance.

Statistical Modeling

In the same analysis, Fisher et al. (2003) measured the "power" (i.e., correct detections of an effect) and false-positive rates of two similar statistical modeling procedures, general linear modeling and interrupted time-series experiment (ITSE) modeling. Results showed the statistical models produced “hits” (i.e., correctly identified true effects) at lower rates compared to the split-middle, dual-criteria, and CDC methods. A critique by Huitema (2004) revealed more problems with the ITSE method. Huitema found two statistical tests that comprise the ITSE, an omnibus *F* test and a *t* test, can produce contradictory findings for the same data set. For example, the *F* test might indicate there was a significant change from baseline to treatment when *t* tests show the data are not significantly different.

Conservative Dual-Criteria Method

Fisher et al. (2003) developed the CDC method to improve the accuracy of previous visual inspection methods. The CDC method projects two lines, one based on level and the other based on trend, from the baseline phase to the treatment phase. The number of data points in the treatment phase that fall above or below the two lines is used to determine whether behavior changed significantly from baseline to treatment. The authors demonstrated the CDC method produced fewer false positives than the split-middle technique and the above-mentioned statistical methods while identifying true effects at acceptable levels for simulated AB graphs with various phase lengths and amounts of autocorrelation. Further, the authors successfully used two similar in-person training procedures to train individual behavior technicians and groups of behavior analysts to implement the CDC method for simulated AB graphs.

Other researchers have validated the CDC method using real data sets. Lanovaz et al. (2017) showed the CDC method correctly ruled out false positives when used to analyze baseline data published in four scientific journals. Wolfe et al. (2018) assessed the validity and accuracy of the CDC method by comparing expert visual interpretations with CDC interpretations of published AB graphs. Results showed high correspondence between the CDC method and expert visual inspectors.

Statement of the Problem

Taken together, these studies show the CDC method is accurate and can be applied to real AB graphs in behavior-analytic practice. However, training practitioners to implement the CDC method using in-person procedures like those Fisher et al. (2003) used might be expensive as well as impractical in some cases. Because companies and organizations that provide behavior-analytic services often cannot bill insurance companies for time spent training employees, economical, remotely administered training procedures might be more viable. Further, some companies do not have opportunities to assemble all staff during business hours as they are providing services to individuals in need of behavior-analytic intervention. Training employees to use the CDC method might be further hampered by the effort required to implement the CDC method (i.e., plotting the criteria lines and manually counting data points). The current studies addressed these issues by evaluating a remotely administered, video-based training package designed to teach graduate students to use an automated Microsoft Excel sheet to implement the CDC method.

Method

Subjects and Setting

Graduate students completing coursework in ABA served as subjects. Seventeen subjects returned signed consent forms. Eight subjects completed baseline: five in Study 1 and three in Study 2. For Study 1, subjects' data were included in the primary analysis if he or she completed at least three posttraining sessions or completed all available graph packets. For Study 2, subjects' data were included in the primary analysis if he or she completed at least three posttraining plus modified instructions sessions. All subjects received extra credit in one or more of their courses for participating. Table 1 shows the subjects' demographic information and level of experience with ABA and visual inspection.

All sessions were completed electronically on the subjects' personal computers. Subjects completed sessions at their own pace. For subjects included in these analyses, total participation time ranged from 4.9 to 19.9 weeks with an average of 14.5 weeks.

Materials

Graph packets. We created 14 graph packets, each comprised of 63 AB graphs with 20 baseline data points and 20 treatment data points. All graphs were comprised of simulated data generated using an equation based on the one Fisher et al. (2003) used. We changed Fisher et al.'s equation to include a variability parameter and negative intervention effects. Our modified equation was:

$$Y_i = 10 + D + (1 - \alpha)(V)E_i + \alpha(V)(E_{i-1})$$

Y_i was the value of the dependent variable at time i, D was the intervention effect, α was the autocorrelation parameter, V was the variability parameter, E_i was the error term at time i, and E_{i-1} was the error term at time i-1. The intervention effect had seven values (0, 1, -1, 3, -3, 5, and -

5), the autocorrelation parameter (*a*) had three values (0, 0.3, and 0.5), and the variability parameter (*V)* had three values (1, 2.5, and 5). Each of the 63 combinations of the three parameters (e.g., D is -1, α is 0.3, and V is 5 is one combination) occurred once per graph packet. All observations had a unique error term randomly selected from a normal distribution with a mean of 0.0 and a standard deviation of 1.0. We changed all negative values for Y_i to zero.

Each subject received one graph packet each session. Each graph packet contained an Excel sheet with 63 graphs and the corresponding raw data as well as a Word document that served as an answer sheet. Each graph was labeled "Reduction" if "treatment" was supposed to decrease behavior (i.e., D was negative) or "Skill Acquisition" if treatment was supposed to increase behavior (i.e., D was positive). Graphs with no programmed treatment effect (i.e., D was zero) were labeled either "Reduction" or "Skill Acquisition" according to a coin flip. Figure 1 shows an example Excel sheet. Each answer sheet included the graph numbers and boxes in which the subjects selected their interpretations as well as the following instructions: "For each graph, mark whether or not the graph ***shows a treatment effect***. If a graph ***does*** show a treatment effect, check "***Yes***." If a graph ***does not show*** a treatment effect, check "***No***." Email me your answer sheet when you are finished."

CDC sheet. In addition to the graph packet (each containing an Excel sheet and an answer sheet), subjects received a copy of the CDC sheet each session. The CDC sheet is a Microsoft Excel workbook developed by Swoboda, Kratochwill, and Levin (2010) that automatically applies the CDC method to different types of graphs. The current study only used the CDC sheet developed for AB graphs. The CDC sheet was comprised of a blank data sheet and a graph with the criteria lines and interpretations. Figure 2 shows an example CDC sheet graph with interpretations.

Response Measurement and Interrater Agreement

The primary dependent variable was agreement—the percentage of subjects' graph interpretations that agreed with the CDC method. Subjects interpreted graphs by checking a box "yes" or "no" on the Answer sheet to show whether each graph showed a treatment effect. We determined agreement with the CDC method by comparing subjects' interpretations with answer keys showing the CDC method's interpretations for each graph in the packet. If an interpretation matched (e.g., both the subject and the answer key indicated graph 42 did not show a treatment effect), we scored an agreement. We measured the subjects' agreement with the CDC method before and after completing the video-based training to assess its effects.

We scored two types of interrater agreement. The first was designed to evaluate the reliability of the answer keys. The first author applied the CDC method for all 14 graph packets, creating the answer keys, and a second rater independently applied the CDC method for three of the 14 graph packets to assess reliability. For each of the three graph packets, we counted the number of interpretations for which the two raters agreed, divided that number by 63 (the number of graphs in each packet), and multiplied by 100% to obtain trial-by-trial agreement. Across the three packets, interrater agreement averaged 99.5% (range, 98.4% to 100%). The second type of interrater agreement assessed the reliability of the subjects' agreement. A second rater independently determined the agreement for at least one-third of sessions for each condition. We used the total count method, dividing the smaller agreement value by the larger one and multiplying by 100%. Interrater agreement averaged 99.5% (range, 96.1% to 100%) for baseline, 100% for posttraining, and 99.6% (range, 98.3% to 100%) for posttraining plus modified instructions.

Accuracy served as a secondary dependent variable. We used the programmed intervention effects (i.e., the value of *D*) to determine the accuracy of subjects' interpretations. For example, if *D* was zero, the correct interpretation would be "no." If *D* was set to any value other than zero, the correct interpretation would be "yes." We scored the proportion of programmed effects incorrectly detected (i.e., false positives) and the proportion of programmed effects correctly detected (i.e., power) for each session.

In addition, we obtained self-reported follow-up data from some subjects at least one month after they completed the study. Subjects submitted follow up data via email by responding either "yes" or "no" to the question: "Have you used the CDC method since finishing the study?"

General Procedure and Design

For both Study 1 and Study 2, subjects received one graph packet and the CDC sheet via email each session, except for video-based training sessions. Subjects received their graph packets in random order. Excluding video-based training sessions, we included the following instructions in the body of each email: "I attached a copy of a graph packet and its answer sheet to this email. For all sessions, your job is to mark on the answer sheet whether each graph in the packet shows a treatment effect. Once you've completed the answer sheet, please save it and email it to me. After I review your completed answer sheet, I'll be able to send you the graph packet and answer sheet for the next session. Be sure to review the instructions on the top of the answer sheet." Subjects had unlimited time to review the graphs and enter their interpretations. Within conditions, subjects progressed to the next session after submitting each answer sheet to the first author. For both studies, we used a multiple baseline across subjects design to show experimental control.

Study 1.

Baseline. In the baseline condition, no programmed consequences were in place for agreement with the CDC method. Subjects proceeded from baseline to training based on stable responding.

Video-based training. After baseline, subjects completed a video-based training similar to that used by Fisher et al. (2003). Subjects received written instructions describing each of the seven steps required to apply the CDC method using the CDC sheet and a 5-min training video. In the video, the first author gave vocal instructions on applying the CDC method with the sheet and modeled its use for two graphs similar to those included in the graph packets. One graph showed a treatment effect but the other did not. To ensure each subject watched the video, he or she sent us an email listing three six-character codes embedded at different points in the video. Subjects who did not provide all three six-character codes or provided incorrect codes would have been excluded from data analysis; this did not occur. After each subject completed the training, he or she began posttraining.

Posttraining. After training, subjects completed treatment sessions identical to baseline. The number of sessions each subject completed was determined by patterns in responding.

Study 2.

Baseline. Baseline sessions were identical to those in Study 1. Subjects progressed from baseline to video-based training contingent on stable responding.

Video-based training. Training sessions were identical to those in Study 1. Again, subjects proceeded to the posttraining condition if and when they submitted the correct codes. Both subjects submitted the correct codes.

Posttraining. Posttraining sessions were identical to those in Study 1. If a subject did not show an immediate increase in agreement in his or her first posttraining session, he or she progressed to the posttraining plus modified instructions condition.

Posttraining plus modified instructions. This condition was designed to separate the effects of the video-based training from each subject's preference for using unassisted visual inspection or the CDC method. In other words, a subject might have learned to use the CDC method after completing the video-based training but continued to use unassisted visual inspection due to personal preference. Thus, low agreement in the posttraining condition could reflect preference for using unassisted visual inspection rather than an ineffective training procedure. This condition included specific instructions to use the CDC method to control for preference.

Posttraining plus modified instructions sessions were identical to posttraining sessions with the exception that an additional, bolded sentence instructing subjects to use the CDC method was added to the instructions in the body of each email. For these sessions, the instructions in the body of each email read: "I attached a copy of a graph packet and its answer sheet to this email. For all sessions, your job is to mark on the answer sheet whether each graph in the packet shows a treatment effect. **Specifically, be sure to use the CDC sheet to interpret each graph using the CDC method.** Once you've completed the answer sheet, please save it and email it to me. After I review your completed answer sheet, I'll be able to send you the graph packet and answer sheet for the next session. Be sure to review the instructions on the top of the answer sheet." The number of posttraining plus modified instructions sessions each subject completed was determined by patterns in responding.

Results

Study 1

Figure 3 shows the percentage of interpretations that agreed with the CDC method across subjects in baseline and posttraining. For S15, agreement was consistently higher in the posttraining condition than in baseline, evidenced by the lack of overlapping data points. For S6, levels of agreement in posttraining were higher than those in baseline for only two of the four data points. In addition, agreement dropped significantly below baseline levels during session 13, showing the video-based training did not consistently improve agreement with the CDC method. Finally, S1 showed a gradual increase in agreement across 13 baseline sessions, suggesting the other subjects' agreement might have increased due to practice effects rather than the effects of the video-based training.

Figure 4 shows the false positives and power of the subjects' interpretations across conditions. Contrary to Matyas and Greenwood's (1990) results, subjects did not show high levels of false positives in any condition, despite our inclusion of graphs with moderate-to-high levels of programmed autocorrelation and variability. For S15, consistently high levels of agreement with the CDC method in the posttraining condition coincided with decreases in both false positives and power. In contrast, S6, whose posttraining data did not suggest consistent, accurate use of the CDC method, showed an increase in false positives following training. These data tentatively suggest consistent and accurate use of the CDC method might result in conservative, but moderately powerful, interpretations of graphed data.

Study 2

Figure 5 shows the percentage of interpretations that agreed with the CDC method across subjects and conditions. For both S17 and S8, levels of agreement did not increase from baseline in either posttraining condition. These results show the video-based training did not effectively increase agreement with the CDC method, even when modified instructions were included to control for preference.

Figures 6 shows the false positives and power of the subjects' interpretations across baseline and posttraining plus modified instructions. For S17, false positives decreased from baseline to posttraining plus modified instructions while power increased. Data for S8 show little change in both false positives and power between conditions. Because one would observe low false positives and moderate power in the posttraining plus modified instructions condition had either subject consistently used the CDC method, these data confirm neither S17 nor S8 consistently used the CDC method following the video-based training.

Discussion

Results from both studies show the video-based training was generally ineffective. Although S15 showed an immediate and consistent increase in agreement following the video-based training, S6, S17, and S8 did not. Further, that none of the subjects in Study 2 showed increases in agreement following training suggests the lack of consistently high levels of agreement following training was due to an ineffective training package rather than preference for other data analysis methods.

One reason the video-based training might have been ineffective is the increased complexity of applying the CDC method in this study compared to previous studies. Previous studies taught subjects to apply the CDC method to graphs that already showed the criteria lines (Fisher et al., 2003; Stewart et al., 2007) whereas the current studies required subjects to generate the criteria lines themselves using the CDC sheet. When using the CDC sheet, subjects had to complete a seven-step behavior chain to correctly apply the CDC method, resulting in seven opportunities for errors. Training packages that incorporate chaining procedures (e.g., forward chaining) might be more effective.

Another interpretation is the video-based training's lack of programmed consequences rendered it ineffective. The current studies' video-based training package was only comprised of antecedents—instructions and modeling. Stewart et al. (2007) effectively taught the CDC method using instructions, modeling, and a quiz. The quiz component required subjects to accurately apply the CDC method before progressing to the next phase. The programmed consequences inherent in the quiz might have improved the efficacy of their training package. Consistent with this interpretation, Ward-Horner and Sturmey (2012) conducted a component analysis designed to identify the most effective components of Behavioral Skills Training (BST), a training package comprised of antecedent and consequent components. The authors found that, when using BST to teach teachers to administer functional analysis conditions, antecedents alone (e.g., modeling) were effective for two teachers whereas a third teacher required a consequence component (i.e., feedback) to reach mastery.

Finally, it is possible antecedent-only training packages can teach people to use the CDC method but that remotely administered antecedents, like those used in the current studies, are too weak to be effective. The results of Fisher et al.'s (2003) in-person training package support this conclusion. Fisher et al.'s training package and the current studies' training package consisted of the same components (instructions and modeling) but only Fisher et al.'s proved effective. However, other factors might be responsible for the difference in effectiveness. For example, it is possible that the trainers in Fisher et al.'s study inadvertently provided some consequences for responding, such as answering subjects' questions while modeling the procedure.

Nevertheless, the CDC method remains a reliable, conservative, and moderately powerful data analysis method. For one graph packet used in the current studies, the CDC method yielded an overall accuracy of 58.7%, a low false-positive rate (11.1%), and relatively high accuracy when the autocorrelation (66.7%) and variability (57.1%) variables were set to their highest parameters. Future research should continue to identify economical, remotely administered training packages designed to teach students and practitioners of ABA to use the CDC method. Specifically, researchers should evaluate the effects of remotely delivered training packages that include consequent components and chaining procedures. Further, few studies have assessed the maintenance of using the CDC method, especially by active ABA practitioners. Research in this vein might identify cost-effective maintenance strategies to ensure practitioners continue to use the CDC method after training ends. Thirdly, researchers should investigate the effects of learning to apply the CDC method on learning to conduct unassisted visual inspection. Because it includes visual aids for level and trend as well as elements of the steady-state strategy, the CDC method might function as a prompt for the component behaviors that make up unassisted visual inspection. Thus, the CDC method might be used to increase the efficiency of teaching students and practitioners traditional, unassisted visual inspection. For example, the CDC method's criteria lines could be used as stimulus prompts in an errorless learning arrangement.

Although the CDC method is empirically supported, it might not be ideal for all situations in behavior-analytic practice. For example, the CDC method might identify a treatment effect in highly variable data whereas an experienced visual inspector might reject a treatment effect due to the variability. This problem is especially relevant to behavior-analytic practice where uncontrolled environments often produce highly variable data. Further, previous research validating the CDC method has used relatively short phases (i.e., three to 20 data points).

Practicing behavior analysts and behavior technicians are often tasked with analyzing data from longer treatment phases. Future research should focus on extending the generality of the CDC method to these situations.

Finally, we should note that the CDC method is not inherently superior to unassisted visual inspection procedures. Indeed, the significance thresholds incorporated in the CDC method are arbitrary. It is certain the CDC method would reject some convincing AB graphs and, therefore, impede clinical progress if used in practice without supervision by a competent behavior analyst. Because it emphasizes level, trend, and baseline logic, is easy to implement, and showed high correspondence with expert interpretations, the CDC method is an ideal first step in learning to conduct visual analysis. We recommend behavior technicians and students seeking advancement in behavior-analytic practice be trained to use the CDC method provided they continue to develop their understanding of behavior analysis and the philosophy of science.

References

Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). Boston: Pearson Education, Inc.

Behavior Analyst Certification Board. (2014). Professional and Ethical Compliance Code for Behavior Analysts. Retrieved from: http://bacb.com/wp-content/uploads/2015/08/150824-compliance-code-english.pdf

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, New Jersey: Pearson Education, Inc.

Fisher, W. W., Kelley, M. E., & Lomas, J. E. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis, 36,* 387-406.

Gentile, J. R., Roden, A. H., & Klein, R. D. (1972). An analysis-of-variance model for the intrasubject replication design. *Journal of Applied Behavior Analysis, 5,* 193-198.

Huitema, B. E. (2004). Analysis of interrupted time-series experiments using ITSE: A critique. *Understanding Statistics*, *3*(1), 27–46.

Kahng, S. W., Chung, K., Gutshall, K., Pitts, S. C., Kao, J., & Girolami, K. (2010). Consistent visual analyses of intrasubject data. *Journal of Applied Behavior Analysis, 43,* 35-45.

Lanovaz, M. J., Huxley, S. C., & Dufour, M. (2017). Using the dual-criteria methods to supplement visual inspection: An analysis of nonsimulated data. *Journal of Applied Behavior Analysis , 50,* 662-667.

Matyas, T. A. & Greenwood, K. M. (1990). Visual analysis of single-case time series: Effects of variability, serial dependence, and magnitude of intervention effects. *Journal of Applied Behavior Analysis, 23,* 341-351.

Michael, J. (1974). Statistical inference for individual organism research: Mixed blessing or curse? *Journal of Applied Behavior Analysis, 7*, 467-653.

Sidman, M. (1960). *Tactics of scientific research*. ISBN-13: 978-0962331107.

Stewart, K. K., Carr, J. E., Brandt, C. W., & McHenry, M. M. (2007). An evaluation of the conservative dual-criterion method for teaching university students to visually inspect AB-design graphs. *Journal of Applied Behavior Analysis, 40,* 713-718.

Swoboda, C. M., Kratochwill, T. R., & Levin, J. R. (2010). Conservative dual-criterion method for single-case research: A guide for visual analysis of AB, ABAB, and multiple-baseline designs (WCER Working Paper No. 2010-13). Retrieved from University of Wisconsin-Madison, Wisconsin Center for Education Research website: http://www.wcer.wisc.edu/publications/workingPapers/papers.php

Ward-Horner, J. & Sturmey, P. (2012). Component analysis of behavior skills training in functional analysis. *Behavioral Interventions, 27*, 75-92.

Wolfe, K., Seaman, M. A., Drasgow, E., & Sherlock, P. (2018). An evaluation of the agreement between the conservative dual-criterion method and expert visual analysis. *Journal of Applied Behavior Analysis, 51,* 345-351.

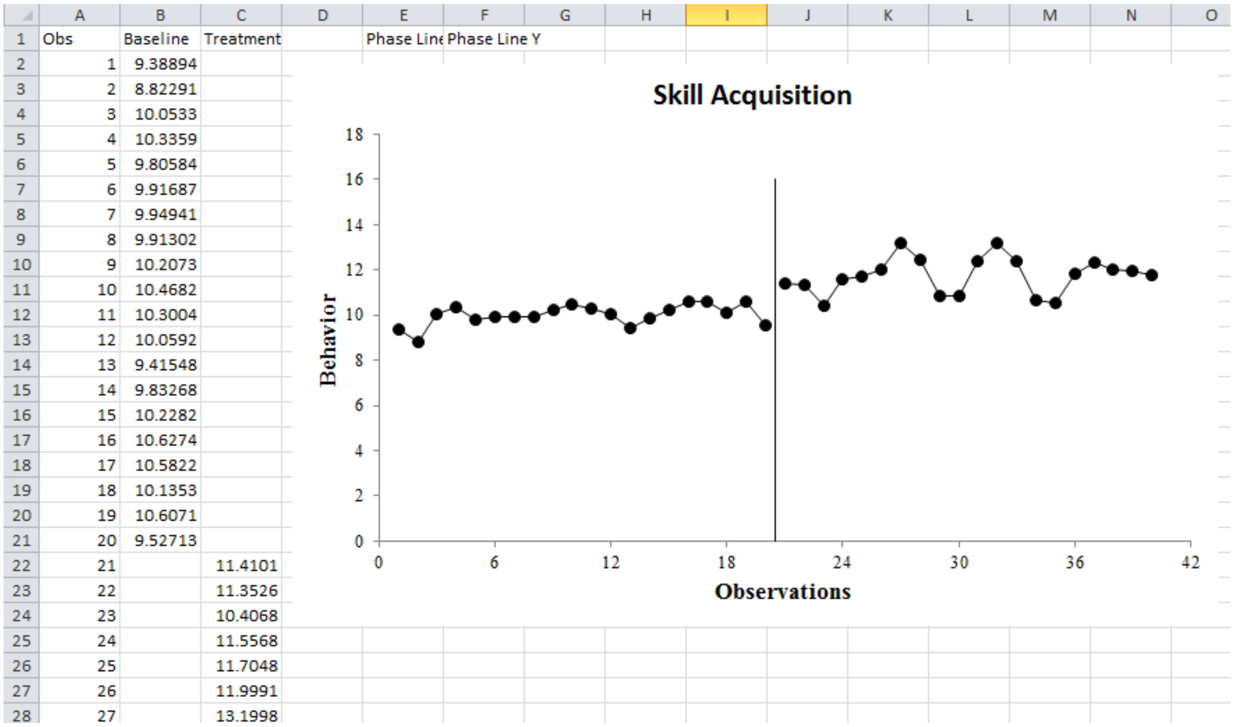
Wolfe, K. & Slocum, T. A. (2015). A comparison of two approaches to training visual analysis of AB graphs. *Journal of Applied Behavior Analysis, 48*, 472-477.

Table 1

Demographic Information

Subject	S1	S6	S8	S15	S17
Age	24	31	26	22	26
Sex	Female	Male	Female	Female	Female
Experience graphing and analyzing data	1 year	0 years	1 year	<1 year	1 year
Experience in ABA	3 years	0 years	2 years	0 years	0 years
ABA credentials	RBT	None	RBT	None	None
Degrees earned	B.A. (unspecified)	B.S. (unspecified)	Bachelor's (unspecified)	B.A. Psychology	B.S. Psychology
Graduate-level instruction completed	3 semesters	0 semesters	3 semesters	0 semesters	3 semesters
Completed any ABA class?	Yes	No	Yes	Yes	Yes

Figure 1. An example raw data sheet and graph.



	A	B	C	D	E	F
1	Obs	Baseline	Treatment		Phase Line	Phase Line Y
2	1	9.38894				
3	2	8.82291				
4	3	10.0533				
5	4	10.3359				
6	5	9.80584				
7	6	9.91687				
8	7	9.94941				
9	8	9.91302				
10	9	10.2073				
11	10	10.4682				
12	11	10.3004				
13	12	10.0592				
14	13	9.41548				
15	14	9.83268				
16	15	10.2282				
17	16	10.6274				
18	17	10.5822				
19	18	10.1353				
20	19	10.6071				
21	20	9.52713				
22	21		11.4101			
23	22		11.3526			
24	23		10.4068			
25	24		11.5568			
26	25		11.7048			
27	26		11.9991			
28	27		13.1998			

Figure 2. An example CDC sheet graph and interpretations for packet 9, graph 15. For this graph, *D* is -5, *α* is 0, and *V* is 2.5. According to the table in the top-left corner, whether the treatment was supposed to increase behavior (i.e., was labeled “Skill Acquisition”) or decrease behavior (i.e., was labeled “Reduction”), this graph does not show a treatment effect.

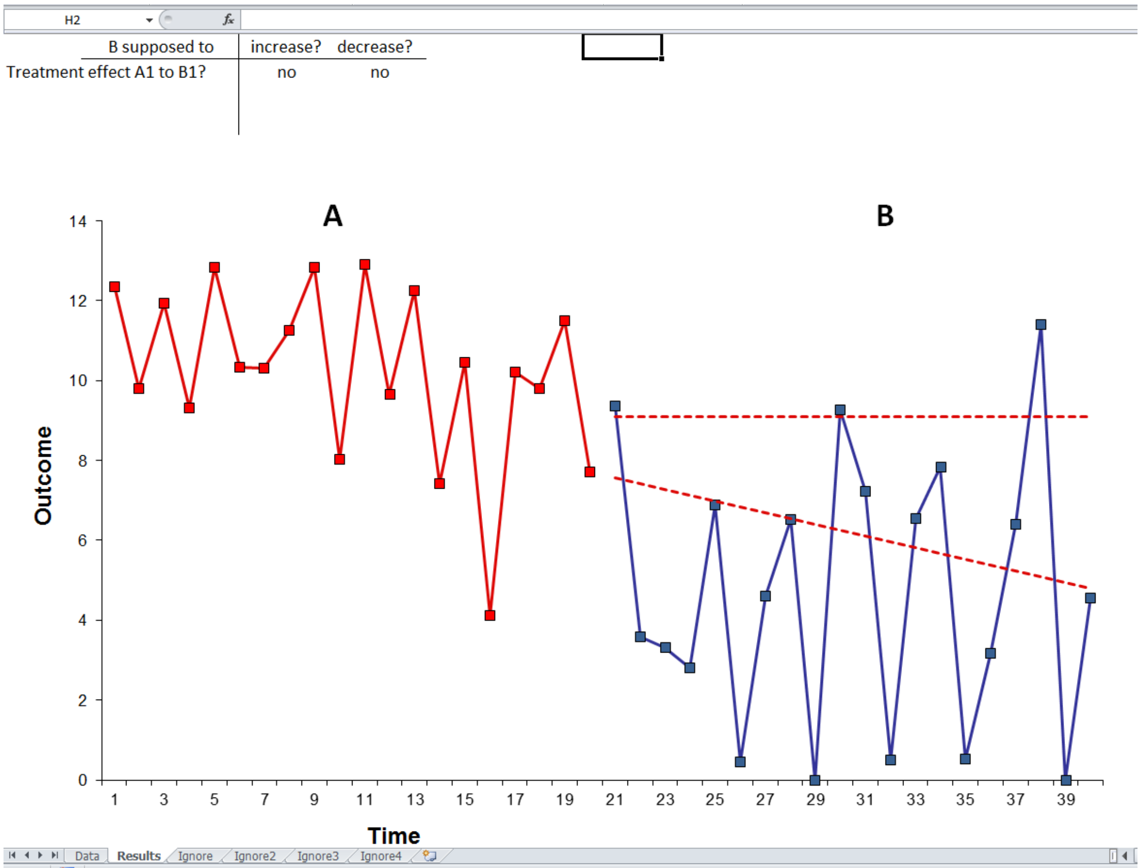


Figure 3. The percent of each subject's interpretations that agreed with the CDC method across sessions for Study 1. Baseline is labeled BL and posttraining is labeled PT.

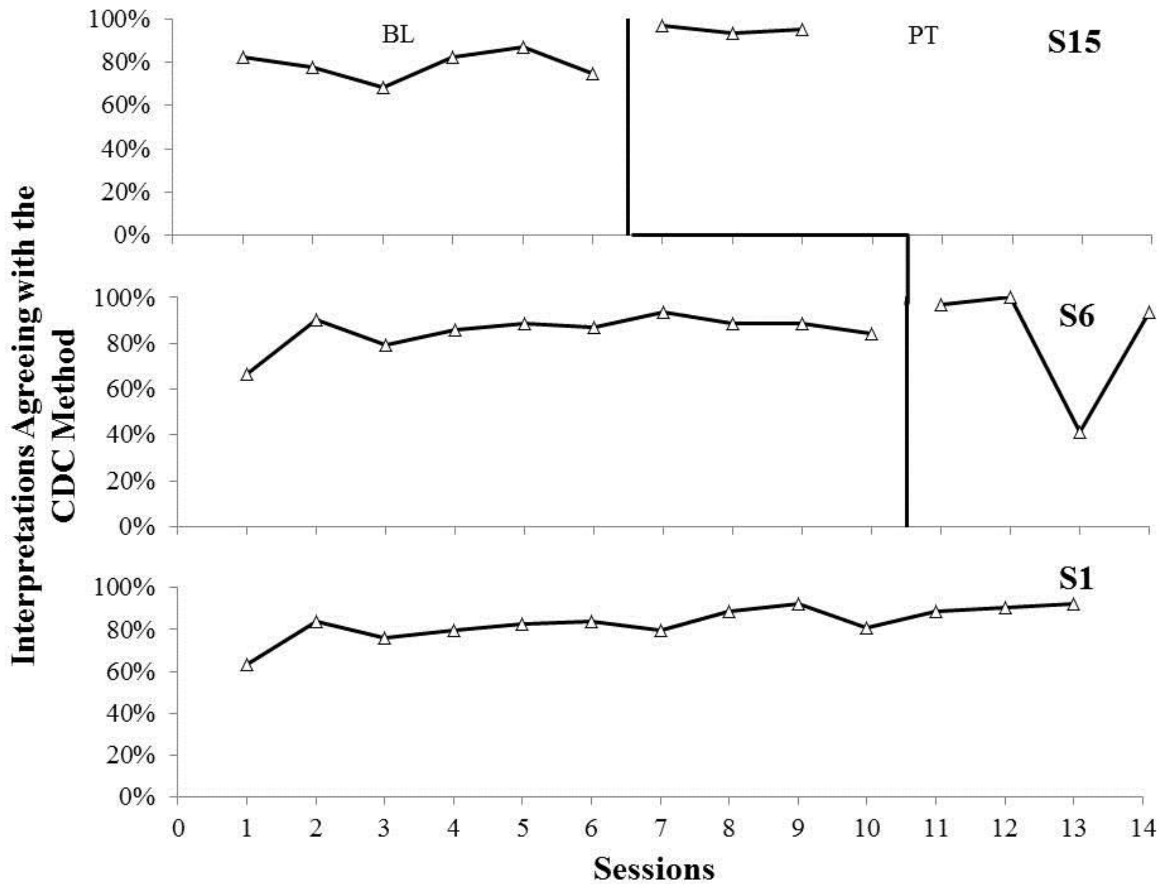


Figure 4. Each subject's average percent of interpretations with a false positive (i.e., Type I error) and average percent of interpretations with a "hit" (i.e., correct detection of a programmed intervention effect) across conditions for Study 1.

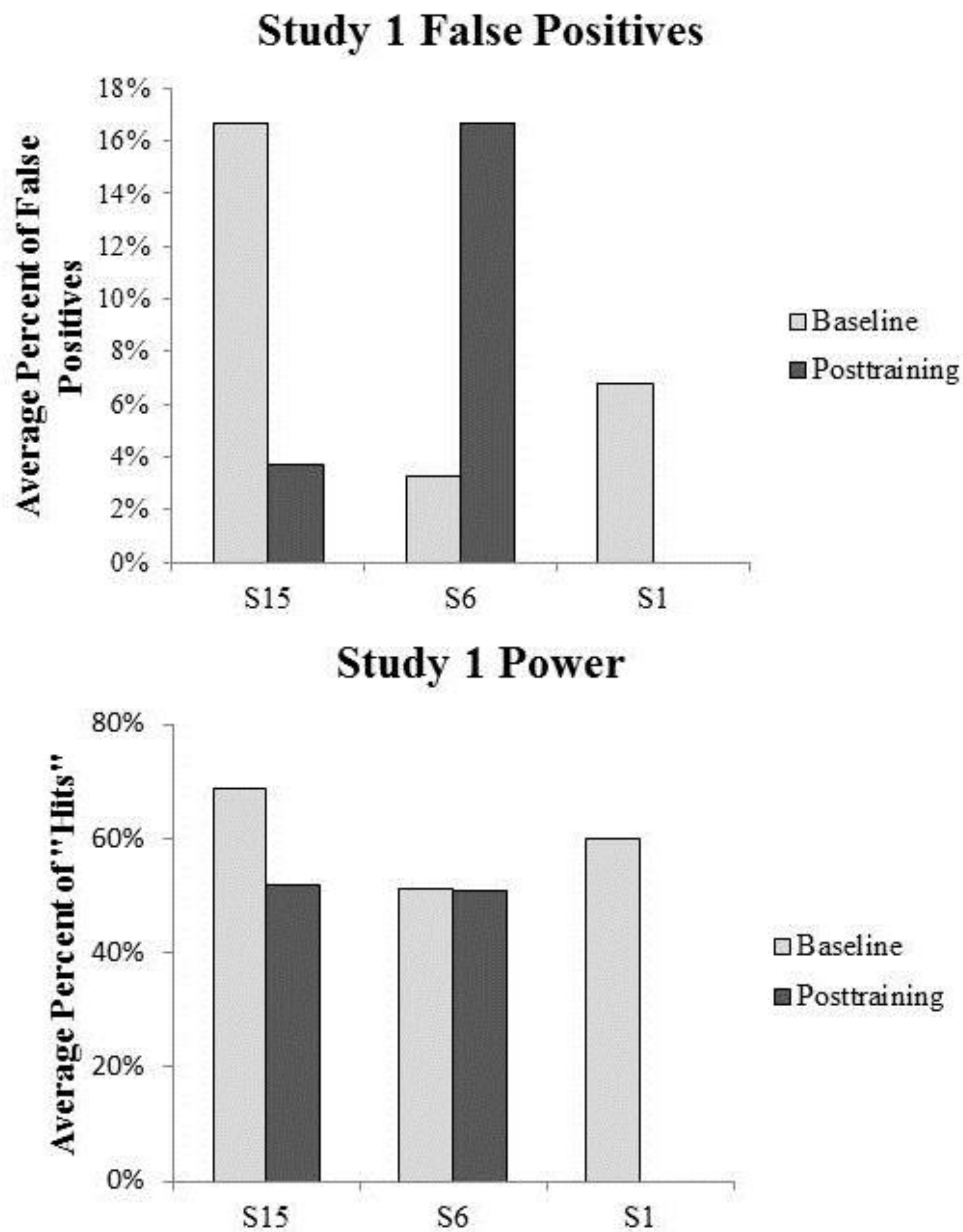


Figure 5. The percent of each subject's interpretations that agreed with the CDC method across sessions for Study 2. Baseline is labeled BL, posttraining is labeled PT, and posttraining plus modified instructions is labeled PT plus MI.

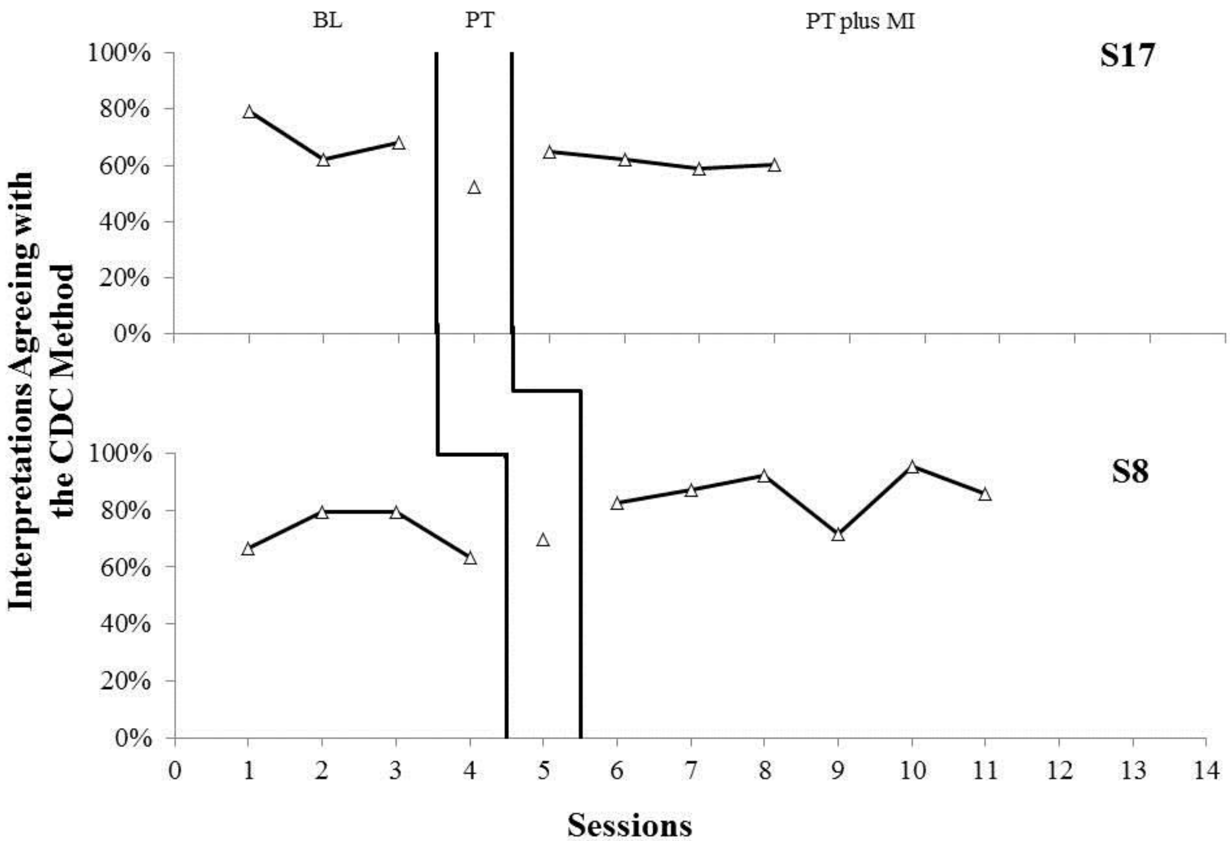


Figure 6. Each subject's average percent of interpretations with a false positive (i.e., Type I error) and average percent of interpretations with a "hit" (i.e., correct detection of a programmed intervention effect) across baseline and posttraining plus modified instructions for Study 2.