

*AN EVALUATION OF THE CONSERVATIVE DUAL-CRITERION
METHOD FOR TEACHING UNIVERSITY STUDENTS TO VISUALLY
INSPECT AB-DESIGN GRAPHS*

KELISE K. STEWART, JAMES E. CARR, CHARLES W. BRANDT, AND
MEADE M. MCHENRY

WESTERN MICHIGAN UNIVERSITY

The present study evaluated the effects of both a traditional lecture and the conservative dual-criterion (CDC) judgment aid on the ability of 6 university students to visually inspect AB-design line graphs. The traditional lecture reliably failed to improve visual inspection accuracy, whereas the CDC method substantially improved the performance of each participant.

DESCRIPTORS: data analysis, visual inspection, conservative dual-criterion method

Research has demonstrated that, despite the ubiquity of visual inspection of single-case design data in behavior analysis, visual inspectors are not always reliable in their judgments (e.g., DeProspero & Cohen, 1979; Ottenbacher, 1990). Researchers have evaluated a variety of visual inspection aids to improve accuracy. One common method is the use of split-middle lines superimposed over data paths (or over subsequent paths) to assist in the estimation of trend (Kazdin, 1982). However, in a recent investigation, Fisher, Kelley, and Lomas (2003) demonstrated that split-middle lines resulted in unacceptably high Type I error rates.

Fisher et al. (2003) developed a new visual inspection aid, the dual criterion (DC) method. When using the DC method, one first calculates the mean line based on baseline data and then superimposes it over the subsequent

data path. Next, a split-middle line is calculated based on baseline data and extended into the subsequent phase. An effect (i.e., a change in data across phases) is said to exist when a prespecified number of data points have fallen above each of the lines according to a binomial equation. Using thousands of Monte Carlo simulations, Fisher et al. demonstrated that the DC method generally resulted in fewer Type I errors and greater power compared to the split-middle technique across a variety of data characteristics (e.g., effect sizes, degrees of autocorrelation). Fisher et al. went a step further and developed the conservative dual-criterion (CDC) method in which the positions of the mean and split-middle lines were raised (for behavioral acquisition graphs) by 0.25 standard deviations (based on baseline data). The CDC method was shown to be a superior visual inspection aid compared to the DC method.

Although a number of studies have examined the reliability of visual inspection or methods for improving visual inspection, surprisingly little research has focused on teaching visual inspection skills. A noteworthy exception is the study by Fisher et al. (2003), in which the authors demonstrated that the DC method could be quickly taught to behavioral clinicians to improve their visual inspection skills. The present study was designed to extend this finding in the following ways. First, the CDC

This study is based on a thesis submitted by the first author, under the supervision of the second author, to the Department of Psychology at Western Michigan University for the MA degree. We thank Wayne Fuqua and Cynthia Pietras for their helpful comments on an earlier version of the manuscript. We also express gratitude to Wayne Fisher for his assistance with graph construction. Kelise Stewart is now affiliated with the Autism Help Center.

Address correspondence to James E. Carr, Department of Psychology, Western Michigan University, 1903 W. Michigan Ave., Kalamazoo, Michigan 49008 (e-mail: jim.carr@wmich.edu).

doi: 10.1901/jaba.2007.713-718

method was employed rather than the DC method. Second, university students were employed rather than clinicians. This extension is important because most behavior analysts presumably acquire visual inspection skills in university courses; however, little is known about the pedagogy or outcomes of those efforts. As such, the third feature of the current study was an initial treatment phase following exposure to a traditional lecture on visual inspection. When the lecture proved to be ineffective, the CDC method was taught to participants. After the CDC method was shown to be effective, the mean and split-middle lines were removed from the graphs in an effort to determine whether successful visual inspection was dependent on them.

METHOD

Participants and Setting

Six undergraduate psychology majors (4 women, 2 men), between the ages of 18 and 20 years, participated in the study. None of them had previously received a lecture on visual data analysis, served as a research assistant, or reported experience with visual data analysis. All sessions took place in two small research rooms on the campus of a large midwestern university.

Materials

Each data point in this study represents a participant's performance visually inspecting eight AB-design line graphs. AB designs were chosen because detecting change between adjacent data paths in an AB design is a requirement for visually inspecting most single-case experimental designs. Approximately 300 graphs were created for the study. Each packet of graphs contained four graphs that depicted behavior change and four that did not. Of the four graphs depicting behavior change (according to CDC criteria), two depicted behavior acquisition and two depicted behavior reduction. Of the four graphs that depicted no behavior change, two of the graphs missed

CDC criteria by one data point, and two of the graphs missed CDC criteria by two data points.

Graphs were created using Microsoft Excel®, Resampling Stats®, and an autoregressive statistical model described by Fisher *et al.* (2003) (see also Matyas & Greenwood, 1990). The model was used to specify features of the baseline and treatment data paths for the resampling program, which then generated multiple AB-design data sets. In the model, the baseline mean was alternately set at 10, 15, and 20, the standard deviation was set at 0.15, the slope was set at 0, the autocorrelation error was set at 0.25, and the effect size ranged from 0 (no effect) to 3.

Dependent Variables and Interobserver Agreement

The primary dependent variable was the percentage of graphs (out of a packet of eight) correctly identified as displaying behavior change or not. Participants completed at least three packets per phase, which were completed within one to two sessions. A second dependent measure was collected to identify patterns of responding across phases. Using a signal-detection classification system, participants' judgments of each of the graphs were coded as hits (true positives), false alarms (false positives), correct rejections (true negatives), or misses (false negatives).

At least 27% of each participant's graph packets were reevaluated by a second independent observer for interobserver agreement calculation. An agreement was defined as both observers transcribing a participant's judgment of a graph with the same outcome. Interobserver agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100%. Mean agreement exceeded 95% for all participants. Interobserver agreement was also assessed on the classification of each participant's judgment of a graph as a hit, false alarm, correct rejection, or miss. Point-by-point agreement was calculated for at least 27% of

graph packets and averaged at least 95% for each participant.

Experimental Design and Procedure

A withdrawal design embedded within a non-concurrent multiple baseline design across participants (in pairs) was used to evaluate the effects of experimental conditions on visual inspection performance. After baseline, phases were changed when data met CDC criteria. Participants did not receive performance feedback at any point during the study.

Baseline. During baseline, participants were provided with sequential graph packets and asked to determine whether each graph demonstrated behavior change without prior training. The graphs did not include any visual inspection lines (e.g., mean, split middle).

Traditional lecture. During the second phase, the experimenter provided narrative instruction on visual inspection via a 12-min videotaped lecture. This lecture incorporated the basic elements of visual data analysis based on the visual data analysis chapter in Cooper, Heron, and Heward (1987). After answering correctly at least 8 of the 10 closed-item quiz questions on the lecture content, participants were provided with sequential graph packets and asked to determine whether each graph demonstrated behavior change. The graphs did not include any visual inspection lines.

CDC method. During the third phase, the experimenter introduced the CDC method to participants with an 8-min videotaped lecture and brief demonstration. After answering correctly four of the four closed-item quiz questions on the CDC method, participants were provided with the binominal table and sequential packets of graphs containing CDC lines and were asked to determine whether each graph demonstrated behavior change.

Independent Variable Integrity

Independent variable integrity was assessed for the traditional lecture and CDC phases. Undergraduate research assistants used check-

lists of the critical components of visual inspection and the CDC method to determine whether each videotaped lecture included the proper content. Each lecture was deemed to have covered all of the critical components. The checklists used to assess independent variable integrity are available from the second author.

RESULTS AND DISCUSSION

The primary findings for each participant are depicted in Figure 1. Each condition produced consistent results across participants. During baseline, mean accuracy for each participant ranged from 54% (P3) to 66% (P4). The traditional lecture resulted in slightly reduced mean accuracies for each participant (range, 47% for P3 to 64% for P5). The introduction of the CDC method resulted in increased accuracy for each participant, eventually producing perfect or near-perfect accuracy in each case. However, 2 participants required retraining (lecture and demonstration) in the CDC method to produce (P3) or maintain (P1) improved accuracy. The withdrawal of the CDC lines and table during the second baseline resulted in reduced accuracy for each participant, with mean accuracies ranging from 56% (P4) to 70% (P6). The reintroduction of the CDC lines and table for P3, P4, P5, and P6 resulted in accuracy comparable to the first CDC condition in each case.

Figure 2 depicts, for each condition, responses classified as hits, correct rejections, false alarms, or misses. During baseline, P1, P2, and P4 tended to overreport behavior change, with most of their responses classified as hits or false alarms. Conversely, P3 and P6 tended to underreport behavior change, with most of their responses classified as correct rejections or misses. Although the traditional lecture did not improve visual inspection, it altered each participant's response pattern by increasing hits and false alarms for most of them. The first implementation of the CDC method increased hits and correct rejections for each participant, with the former still

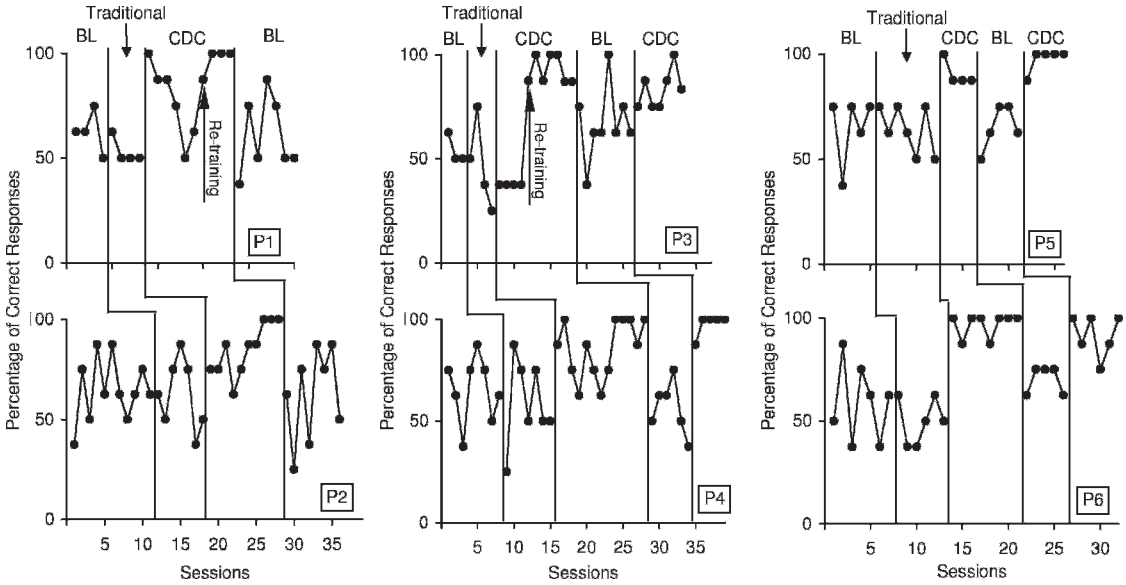


Figure 1. The percentage of correct responses across baseline, traditional lecture, and CDC conditions for each participant.

more common than the latter. The withdrawal of the CDC lines and table resulted in more false alarms for each participant. Finally, the reintroduction of the CDC lines and table for P3, P4, P5, and P6 resulted in increases in both hits and correct rejections.

The present study showed that (a) a traditional lecture on visual inspection was ineffective in improving visual inspection by university students, (b) the lecture resulted in participants consistently overreporting behavior change, (c) the CDC method resulted in substantial improvements in visual inspection, and (d) inspection performance returned to baseline levels when CDC lines were removed from graphs. These findings, however, should be interpreted in the context of at least two limitations. First, all participants received a lecture on visual inspection before being taught the CDC method. It is possible that this history enhanced the effects of the CDC method, constituting a sequence effect. This possibility should be explored in future research. Second, participants were evaluated nonconcurrently, rather than concurrently, in a multiple baseline design. The nonconcurrent

multiple baseline design does not control for history threats to the independent variable, as does the concurrent multiple baseline design. However, the immediacy and reliability of the effects demonstrated across participants appear to mitigate this concern.

With regard to the lecture delivered in the present study, it is possible that the content or pedagogy might have been insufficient to effectively teach accurate visual inspection. Future studies might investigate the effects of additional models and rehearsal opportunities within the lecture method. Furthermore, feedback could be used during rehearsal opportunities to enhance performance. If such an approach fails to improve visual inspection, a more detailed analysis of instructional content would be warranted.

Failure of the CDC method to maintain accurate visual inspection suggests that the method might enhance visual inspection only when the inspection lines and binomial table are present. Unfortunately, this does not bode well for teaching visual inspection to students who will later apply those skills when consuming the published literature, an activity not well suited

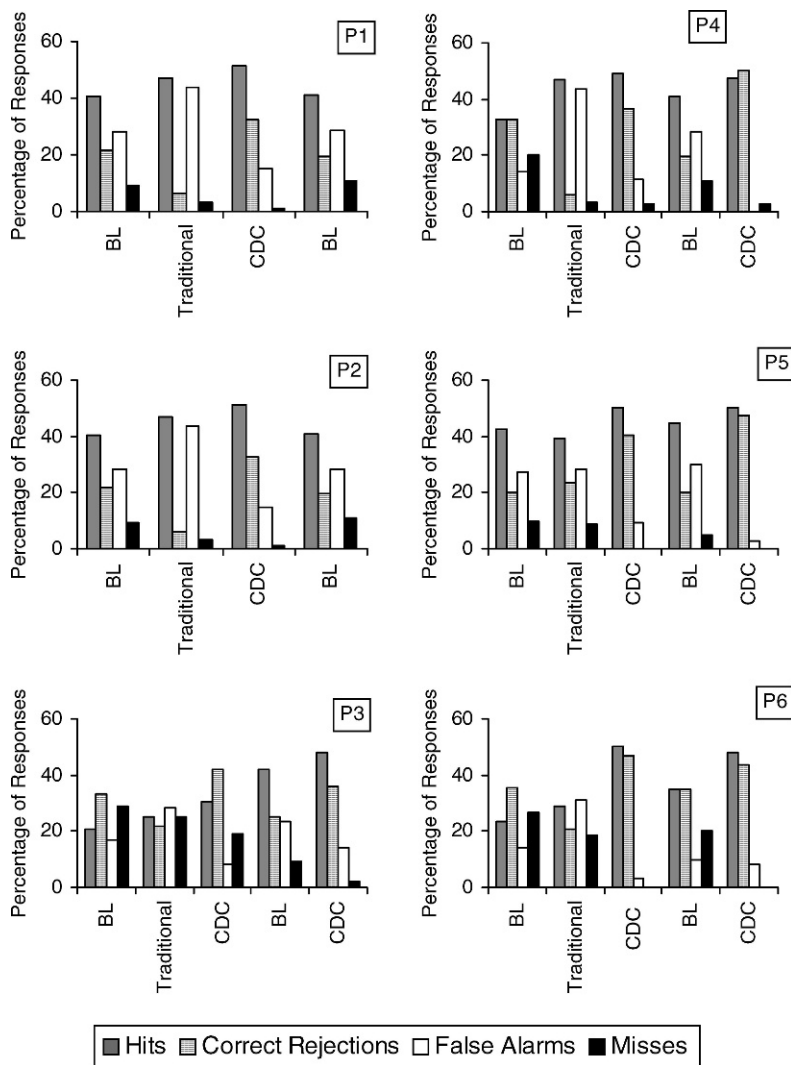


Figure 2. The percentage of responses in each signal-detection category across baseline, traditional lecture, and CDC conditions for each participant.

to the trappings of the CDC method. However, it is possible that a longer history with the CDC method would have resulted in better maintenance when it was no longer available. This possibility warrants additional research. If such research indicates otherwise, it might be necessary to teach participants to visually estimate the CDC lines and to use a verbal heuristic regarding the number of data points required for an effect. If visual inspection performance still is not maintained, research on visual inspection techniques that do not

require the long-term presence of supplemental stimuli would be warranted.

REFERENCES

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis*. Columbus, OH: Merrill.
 DeProspero, A., & Cohen, S. (1979). Inconsistent visual analyses of intrasubject data. *Journal of Applied Behavior Analysis, 12*, 573-579.
 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.

- Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings*. New York: Oxford University Press.
- 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.
- Ottenbacher, K. J. (1990). Visual inspection of single-subject data: An empirical analysis. *Mental Retardation*, *28*, 283–290.

Received July 12, 2006

Final acceptance December 15, 2006

Action Editor, Mark Dixon