1-2011

Evaluating the Effects of a Job-Aid for Teaching Visual Inspection Skills to University Students

Candice M. Jostad

*Western Michigan University*

Follow this and additional works at: http://scholarworks.wmich.edu/dissertations

Part of the Cognition and Perception Commons, Cognitive Psychology Commons, and the Educational Psychology Commons

Recommended Citation


http://scholarworks.wmich.edu/dissertations/423

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
EVALUATING THE EFFECTS OF A JOB-AID FOR TEACHING VISUAL INSPECTION SKILLS TO UNIVERSITY STUDENTS

by

Candice M. Jostad

A Dissertation
Submitted to the
Faculty of the Graduate College
in partial fulfillment of the requirements for the
Degree of Doctor of Philosophy
Department of Psychology
Advisor: James E. Carr, Ph.D.

Western Michigan University
Kalamazoo, Michigan
June 2011
EVALUATING THE EFFECTS OF A JOB-AID FOR TEACHING VISUAL INSPECTION SKILLS TO UNIVERSITY STUDENTS

Candice M. Jostad, Ph.D.
Western Michigan University, 2011

Visual inspection is the primary method of data analysis used in behavior analysis. Thus, it is important that behavior analysts have the skills necessary for accurate visual inspection. Research has shown that visual inspection can sometimes be unreliable, which has broad implications for the evaluation of treatment effects using this method. Traditional lectures have been shown to be ineffective in teaching visual inspection skills to a satisfactory level, although improvements in visual inspection have been accomplished using statistical methods and aids such as celeration lines superimposed on graphs. However, these methods are not effective when the aids are removed and are typically unavailable when inspectors evaluate graphs in natural settings. Experiment 1 of the current investigation evaluated the effects of a portable job-aid on the visual inspection of graphs by university students and found positive results. Experiment 2 assessed the job-aid in a university setting and compared it to traditional lecture. Results showed main effects of both job-aid and traditional lecture, but no significant differences between group means. However, the number of students meeting criterion (i.e., 80% or better) following the job-aid plus teaching package was more than twice the number reaching criterion
following traditional lecture (16 vs. 7). The current research demonstrates a visual inspection tool for which training is brief, it is easy to use, it produces quick and clinically significant results, it is portable, and it is effective in group-instruction circumstances.
ACKNOWLEDGMENTS

I would like to offer my sincerest appreciation to the following individuals for the support they have provided. Without their guidance and encouragement, this project would not have been realized. First, I would like to thank my advisor, Dr. Jim Carr. His unobtrusive but ever-present advisement fostered independence while providing a constant model of professionalism, leadership, and scholarship. Second, I would like to thank the members of my committee, Drs. Wayne Fuqua, Dick Malott, and John Austin, whose sponsorship and direction have shaped both this project and my scientific behavior. Third, I must express my gratitude toward the talented and dedicated researchers with whom I collaborated on these studies: Kerry Ann Conde, Dr. Amanda Karsten, Allison Castile, and Carly Cornelius. This project would not have been completed without their contributions. Fourth, I would like to thank my family, whose names I borrowed for participant pseudonyms, and without whose support I would not have survived graduate school. Finally, I want to offer my deepest thanks to my loving and patient fiancé, Eric, who has tolerated my monopolization of the office in our new home, my neglect of our wedding plans, and countless evenings and weekends without company as I completed this project.

Candice M. Jostad
## TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................ ii

LIST OF TABLES ................................................................................................................ vi

LIST OF FIGURES .............................................................................................................. vii

INTRODUCTION

<table>
<thead>
<tr>
<th>Purpose and Benefits of Visual Analysis</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Reliability of Visual Inspection</td>
<td>6</td>
</tr>
<tr>
<td>Research on Improving Visual Inspection</td>
<td>12</td>
</tr>
<tr>
<td>Structured criteria</td>
<td>12</td>
</tr>
<tr>
<td>Inspection aids</td>
<td>13</td>
</tr>
<tr>
<td>Teaching visual inspection</td>
<td>18</td>
</tr>
</tbody>
</table>

EXPERIMENT I: DEVELOPMENT AND VALIDATION OF THE JOB-AID...... 20

<table>
<thead>
<tr>
<th>Method</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>20</td>
</tr>
<tr>
<td>Setting</td>
<td>21</td>
</tr>
<tr>
<td>Materials</td>
<td>21</td>
</tr>
<tr>
<td>Graphs</td>
<td>21</td>
</tr>
<tr>
<td>Job-aid</td>
<td>22</td>
</tr>
<tr>
<td>Computer teaching presentation</td>
<td>24</td>
</tr>
<tr>
<td>Procedure</td>
<td>24</td>
</tr>
<tr>
<td>Baseline</td>
<td>24</td>
</tr>
<tr>
<td>Job-aid only</td>
<td>26</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Job-aid plus teaching</td>
<td>26</td>
</tr>
<tr>
<td>Job-aid fading</td>
<td>26</td>
</tr>
<tr>
<td>Error correction</td>
<td>27</td>
</tr>
<tr>
<td>Modified job-aid</td>
<td>27</td>
</tr>
<tr>
<td>Trend training</td>
<td>28</td>
</tr>
<tr>
<td>Feedback</td>
<td>28</td>
</tr>
<tr>
<td>Results</td>
<td>29</td>
</tr>
<tr>
<td>Discussion</td>
<td>36</td>
</tr>
<tr>
<td>EXPERIMENT II: EVALUATION OF THE JOB-AID WITH TRAINING IN GROUPS</td>
<td>38</td>
</tr>
<tr>
<td>Method</td>
<td>38</td>
</tr>
<tr>
<td>Participants</td>
<td>38</td>
</tr>
<tr>
<td>Setting</td>
<td>40</td>
</tr>
<tr>
<td>Power analysis</td>
<td>40</td>
</tr>
<tr>
<td>Experimental design</td>
<td>40</td>
</tr>
<tr>
<td>Materials</td>
<td>41</td>
</tr>
<tr>
<td>Graphs</td>
<td>41</td>
</tr>
<tr>
<td>Traditional lecture</td>
<td>41</td>
</tr>
<tr>
<td>Job-aid</td>
<td>41</td>
</tr>
<tr>
<td>Computer teaching presentation</td>
<td>42</td>
</tr>
<tr>
<td>Concept quiz</td>
<td>42</td>
</tr>
<tr>
<td>Dependent measures</td>
<td>42</td>
</tr>
<tr>
<td>Procedure</td>
<td>42</td>
</tr>
<tr>
<td>Pretest</td>
<td>43</td>
</tr>
<tr>
<td>Traditional lecture</td>
<td>43</td>
</tr>
</tbody>
</table>
LIST OF TABLES

1. Summary of statistics for the 10 types of graphs used in the experiment................................................................. 25
LIST OF FIGURES

1. Sample graph with conservative dual criterion (CDC) lines drawn in the second phase .......................................................... 17

2. Sample graph presented to participants ........................................ 23

3. The percentage of accurate visual inspection in baseline, job-aid only, and job-aid plus teaching conditions for Eric (top panel), Dusty (middle panel), and Josh (bottom panel) .............. 31

4. The percentage of accurate visual inspection in baseline, job-aid only, and job-aid plus teaching conditions for Cindy (top panel) and Emily (bottom panel) .................................................. 32

5. The percentage of accurate visual inspection in baseline, job-aid only, job-aid plus teaching conditions, and additional manipulations for Aaron (top panel), and Rylan (bottom panel) .... 33

6. Flowchart depicting the sequence of study procedures ............... 45

7. Group means for participants in Group B (top panel) and Group A (bottom panel) ................................................................. 49

8. Data for individual participants who received feedback in Group B (top panel) and Group A (bottom panel) .......................... 50

9. The percentage of participants in each group scoring 80% or higher at each test ........................................................................... 51
INTRODUCTION

Purpose and Benefits of Visual Analysis

Single-case experimental designs constitute the fundamental analytic methodology in behavior analysis, and the means by which data are evaluated in these designs is through visual analysis of graphic data. Several benefits of this practice are evident to scientists who use it and the consumers of the research produced. Foremost, graphs facilitate a meaningful reaction to the data (Johnston & Pennypacker, 2009). Graphs display relations between behavior and environmental variables in an accessible format, making it easier to identify changes in behavior than by evaluating raw numbers alone (Cooper, Heron, & Heward, 2007; Parsonson & Baer, 1992). Not only do graphs make the data digestible, they also provide detail about the behavior under study (e.g., frequency, duration), the conditions under which it takes place, and environmental variables that influence its occurrence (Parsonson & Baer, 1978).

Graphs are persuasive, durable products of empirical exploration that serve as mediators of scientific discussion (Smith, Best, Stubbs, Archibald, & Roberson-Nay, 2002). A noteworthy benefit of graphed data is their availability to consumers of the research. Data can be presented in a concise form such that little information is lost. Thus, members of the audience may evaluate and interpret the behavior change and its purported causes themselves (Parsonson & Baer, 1978; 1986; 1992). This leads to transparency of the research process,
which contributes to the evolution of the field and the refinement of technologies derived from scientific study.

Furthermore, graphs can be used with clients as a form of feedback on their behavior. Self-monitoring including graphing has been used as a treatment component in successful interventions for a variety of targets, such as increasing the number of steps taken daily by overweight adults (VanWormer, 2004), reducing household energy use (Winett, Neale, & Grier, 1978), increasing the number of pool lengths completed by young swimmers (Critchfield, 1999), reducing excessive caffeine intake (Foxx & Rubinoff, 1979), improving teaching procedures (Kissel, Whitman, & Reid, 1983), and increasing staff compliance with client care routines (Burgio et al., 1990).

Frequent, repeated measurement of the behavior under study is the distinguishing feature of single-case methodology. The immediate graphing that accompanies these recurring observations results in a current visual accounting of behavior that permits flexibility in research design and allows interventions to be easily responsive to the behavior (Johnston & Pennypacker, 2009; Parsonson & Baer, 1978; 1986). It is in this way that behavior-analytic interventions are tailored to meet the needs of individuals (Parsonson & Baer, 1986). The adaptability granted by such intimacy with the data also allows the examination of noteworthy or unusual behavioral variations; if the data take off in an interesting direction, the investigator has the latitude to follow where they lead (Michael, 1974; Parsonson, 2003; Parsonson & Baer, 1978). Furthermore, the
analysis of repeated measures of behavior can provide valuable information about the independent variable, such as whether it has a delayed or temporary effect (Parsonson, 2003). When the data are displayed in graphic form, such changes within phases are readily identified. In contrast, when data are collapsed into averages, this and other potentially useful information is masked.

Indeed, a number of advantages of visual inspection of graphed behavioral data over statistical analysis are apparent. First, and of fundamental importance, statistics and other ways of transforming the data result in abbreviations; some information is necessarily left out (Michael, 1974; Parsonson & Baer, 1978; Parsonson & Baer, 1992). Because of the way the data are transformed in statistical analyses, there is a real possibility that relevant variables may go unexamined. Thus, it follows that if behavior analysts are able to react effectively to their data without such abbreviations, it would be desirable to do so.

Second, in applied behavior analysis, clinical and social significance are valued above statistical significance (e.g., Parsonson, 2003). It is possible for problem behavior to be reduced by a statistically significant amount but still occur at a frequency or intensity that continues to interfere with a client’s functioning or that is unacceptable to the individuals in a client’s life. To the behavior analysts working with that client, despite statistical significance, the change is insufficient to warrant a claim of success. Instead, behavior analysts continue to evaluate treatments that may further reduce problem behavior to
manageable levels. Visual inspection lends well to this focus on clinically meaningful results because it is designed to identify clear and robust effects. However, the use of statistical analysis is beneficial in circumstances in which small effects are expected or when small effects are important.

Third, visual inspection is a more conservative approach than statistical significance testing (Parsonson & Baer, 1986). Whereas practitioners using statistical analyses typically set the acceptable rate of Type I error at .05 (in essence, allowing 5% of findings deemed significant to be attributable to chance), those using visual inspection prefer to “miss” some interventions that might cause weak or subtle effects (i.e., make more Type II errors) than to conclude erroneously that inert interventions are effective (i.e., make Type I errors; Parsonson & Baer, 1986). Replication of effects, which is a component of single-case experimental design, further reduces the likelihood of Type I errors (Parsonson & Baer, 1978). These replications are emphasized in the process of visual inspection (Parsonson, 2003). Therefore, through visual inspection, behavior analysts select only interventions that produce strong and generalized effects. This selectivity is purposeful; the field actively avoids detecting weak effects (Baer, 1977). A powerful demonstration of experimental control, including at least one replication, is needed to convince the consumers of research in behavior analysis (Sidman, 1960). The practice of harvesting only the most potent variables has resulted in swift progression of the field of behavior analysis (Parsonson & Baer, 1986). The discovery of strong, dependable variables has
promoted application in the real world because these variables were likely to persist in their effects even in imperfect conditions (Parsonson & Baer, 1978).

Fourth, unlike statistical analysis, visual inspection does not rely on arbitrary standards for determining the significance of results, nor do the data have to meet certain assumptions for visual inspection to be used (Parsonson, 2003). Because visually inspected data do not have to conform to test assumptions, research designs are able to be guided more by the data than by the analyses intended for those data (Michael, 1974).

Finally, tighter experimental control is encouraged when data are evaluated by visual inspection (Sidman, 1960). In single-case design, variability is more likely to be identified and examined rather than collapsed into averages or “controlled” by statistical means (Michael, 1974). This may lead to the investigation of additional variables, enhancing the breadth of knowledge in our science.

In research as we expand our understanding of the principles of our science and refine the technologies that have derived from them, and in practice as we apply what we have learned to make meaningful changes in the lives of our clients, behavior analysts rely on single-case designs and visual analysis of data to make judgments about the utility of interventions they assess. Visual inspection of graphed data is the primary means of analysis in the field, and is widely accepted by behavior-analytic researchers as the most appropriate method of analyzing within-subject data. However, studies examining visual
inspection have found it to be unreliable in some circumstances. These studies warrant further discussion.

**The Reliability of Visual Inspection**

DeProsero and Cohen (1979) created ABAB-design graphs that demonstrated four visual data patterns thought to affect visual inspection: mean shift pattern, mean shift magnitude, variability, and trend/slope. Variations of each factor were depicted in all data sets. For example, one graph might have depicted an ideal pattern of mean shift (i.e., changes in each B phase in the direction expected by treatment and reversal to baseline levels in the second A phase), a moderate magnitude of mean shift, high variability in the data, and a slope of zero in all phases. Experienced behavior analysts rated the degree to which they were convinced of the functional control depicted in each graph and listed the features of each graph that influenced their decisions. The authors found low interrater agreement among the 108 respondents, with an average correlation of .61 between raters judging the same graph. Furthermore, the factors reported by the raters to have influenced their judgments did so through interactions, rather than singly. That is, mean shift pattern or magnitude, variability, or trend alone did not determine the raters’ confidence in the demonstration of functional control depicted in the graphs; rather, manipulation of these factors combined to produce their effects. This finding reflects the complexity of visual data analysis, which may contribute to the low rates of interrater reliability found between even experienced inspectors.
Gibson and Ottenbacher (1988) reported similarly low levels of agreement among a sample of 20 occupational and physical therapists. The authors examined the four factors described by DeProspero and Cohen (1979), and manipulated two additional factors in their graphs that might influence visual inspection: overlap and serial dependency (i.e., correlation between data points in a series). The therapists, who all had professional experience with individuals with developmental disabilities, rated 24 AB-design graphs on the degree to which they agreed that a significant change was demonstrated across phases. Consistent with previous studies, interrater reliability was low, with coefficients of .52 to .66. Furthermore, the authors confirmed the relations found by DeProspero and Cohen (1979) between graph characteristics (e.g., degree of mean shift, changes in slope) and interrater reliability. They also noted that these factors had considerable effects on rater confidence in their decisions.

More recently, Danov and Symons (2008) compiled packets of multielement graphs depicting real but decontextualized functional analysis data from articles published in a prominent peer-reviewed journal. Forty-three graduate students and faculty members from graduate programs accredited by the Association for Behavior Analysis International assigned functional categories to 26 graphs; these categorizations were then compared to those of the authors who had published each data set. The authors determined the degree of agreement between raters overall, the consistency of individual raters across graphs, and the level of agreement for each graph across all raters. All possible
pairs of raters were assembled and their data compared to determine an overall correlation between raters of .63. The mean consistency of individual raters (i.e., the percentage of graphs with which the rater’s classification matched the published classification) was 69%. Finally, the average percentage of rater agreement with published functional classification was 83% for profiles depicting only one behavioral function and 62% for profiles for which multiple behavioral functions were identified. Importantly, this study included an assessment of the reliability of visual inspection with data presented in multielement design graphs, an experimental design frequently used in applied behavior analysis, particularly with functional analysis. The results suggest that the reliability of visual inspection for these graphs is comparable to that for other single-case design graphs.

Although research has shown that visual inspection can sometimes be unreliable, the discipline of behavior analysis has nevertheless progressed. Membership in professional organizations is growing, submissions to major journals are increasing, applied behavior analysts are in high demand, and the science continues to advance. Indeed, a preponderance of the evidence (i.e., the history of success of the fields of the experimental analysis of behavior and applied behavior analysis) shows that the methods are sufficiently dependable (e.g., Johnston & Pennypacker, 2009; Michael, 1974; Parsonson & Baer, 1978). Therefore, it is important to reconcile this reality with the findings of studies demonstrating the lack of reliability of visual inspection.
One possible explanation is that the investigations are flawed. To be sure, the methods used in such studies have come under scrutiny (e.g., Fisher, Kelley, & Lomas, 2003; Hagopian et al., 1997; Parsonson, 2003; Parsonson & Baer, 1992). To start, studies assessing the reliability of visual data analysis generally are not conducted under conditions in which visual inspection actually occurs (Parsonson). For example, background information is typically removed from the graphs to be evaluated, including information about the participant and the behavior under study, the independent variable characteristics, the setting, and other means to contextualize the data (e.g., DeProspero & Cohen, 1979; Gibson & Ottenbacher, 1988; Normand & Bailey, 2006). Additionally, completed graphs are often used (i.e., graphs of assessments or interventions that have been finished), and this is not how visual inspection progresses in the real world. Instead, behavior analysts react to data on an ongoing basis as they are collected (Cooper et al., 2007; Parsonson & Baer, 1992). Further problems with these studies include that simulated rather than actual data are often used, data are frequently presented in A-B designs, within-phase changes are customarily ignored, and the experience level of judges varies widely across studies (Normand & Bailey; Parsonson; Parsonson & Baer, 1992).

Finally, a considerable limitation is that there is “no known truth” in many studies of the reliability of visual inspection (Parsonson & Baer, 1992, p. 37). In some cases, visual inspection judgments are compared between all raters; in some, results from novice participants are compared with the opinions of expert
participants; in some, judgments are compared to the original authors’ findings (when the graphs are taken from published work); and in others, visual inspection judgments are compared to the findings of statistical tests of significance. Much debate could be spurred by claiming that any one of these sources represents the “truth” (Parsonson). Taken together, the limitations of prior studies suggest that all visual inspection may not be as unreliable as these studies indicate. In fact, a more recent investigation by Kahng et al. (2010) found a high degree of correlation between well-trained, experienced analysts.

Kahng et al. (2010) replicated the prominent study by DeProspero and Cohen (1979) by asking experienced visual inspectors to rate their degree of satisfaction with the functional control demonstrated in each graph on a scale of 0 to 100. The set of 36 graphs differed along four dimensions within and/or across phases (pattern of mean shift, magnitude of mean shift, variability, and trend). Notably, the researchers sampled only individuals who were members of the editorial board or who had served as associate editors of the field’s flagship applied journal over a two-year span. Additionally, Kahng et al. extended DeProspero and Cohen’s investigation by asking raters to provide an answer on a dichotomous scale regarding whether a functional relation (for which the authors provided a definition) between the independent and dependent variables was demonstrated in each graph. Kahng et al. reported a Pearson correlation coefficient for the 100-point scale among all rater pairs of .93, substantially greater than the .61 found by DeProspero and Cohen more than 30 years earlier.
For the dichotomous measure, the participants’ judgments produced a kappa of .84, a desirable level of interrater agreement. The authors concluded that high levels of agreement were possible when highly trained and experienced raters conducted visual analysis.

Even if visual inspection were as unreliable as earlier studies suggest, as Sidman (1960) stated, science is self-correcting. Through replication, further exploration, and application, researchers uncover problematic findings. Thus, even if visual inspection is not always highly reliable, faulty conclusions are identified when they do not withstand further testing. It is possible that behavior analysts’ emphasis on replication has therefore contributed greatly to the progression of the science, in that if erroneous conclusions have been reached based on inaccurate visual inspection, the errors are likely to be discovered when the experimental results in question cannot be reproduced. However, although such potential errors may be corrected over time, there remain serious consequences of unreliable visual inspection.

Clinically, the negative consequences of inaccurate judgments of treatment effects or non-effects are (a) the potential rejection of interventions that may be truly beneficial (a Type II error), likely leading to the termination of treatments that may actually help clients; and (b) the acceptance of treatments that may have no genuine therapeutic effects (a Type I error), therefore wasting the time of therapists and clients. In the research setting, the negative consequences of inaccurate judgments of independent variable effects include
(a) the premature abandonment of potentially fruitful lines of research, which
could change the course of research in the field and delay the development of
techniques that might be useful; and (b) the continuance of lines of research on
variables that do not result in systematic change, wasting valuable experimental
time and resources. Therefore, as a field driven by data, it is essential that we
have the tools to evaluate those data accurately.

Research on Improving Visual Inspection

Structured criteria. Hagopian et al. (1997) developed and evaluated
structured criteria to assist in the visual inspection of multielement design graphs
depicting data from functional analyses. Initially, the authors assessed the visual
inspection performance of three predoctoral interns and found low agreement on
their interpretations of functional analysis data. Next, functional analysis experts
evaluated 64 graphs and described the factors on which they based their
decisions. The variables influencing the experts’ visual inspection of the
multielement design graphs were used to formulate a set of criteria to guide the
inspection of functional analysis graphs. These criteria were to be applied to
multielement design graphs with 10 data points per condition. First, the
individual evaluating the graph considered the data from the control condition
and drew criterion lines at one standard deviation above (upper criterion line)
and below (lower criterion line) the mean of the data in that condition. The
inspector then evaluated the data from each test condition. The function tested
in a condition was determined to be a maintaining variable for the behavior if
five more data points fell above the upper criterion line than below the lower criterion line in that condition. Finally, the individual inspecting the data applied decision rules for automatic reinforcement, trends, low-rate behavior, low magnitude of effects, and multiply controlled behavior, if these variables were relevant to the data. After these criteria were developed, they were used to reevaluate the graphs. When following the criteria resulted in decisions contradicting those of the experts, the disagreements were discussed and the criteria were modified to accommodate the exceptions identified. The outcome was a collection of rules designed to facilitate the interpretation of functional analysis graphs. Application of the final structured criteria resulted in agreement with the functions identified by the expert consensus for 94% of the graphs. Finally, the interns who participated in the first phase of the investigation were taught to use the criteria and applied them to additional graphs. Following training, the average agreement between the raters and the investigators improved from 54% to 90%. Importantly, the authors demonstrated that written instructions could be used to improve the reliability of visual inspection of multielement functional analysis graphs. However, the criteria developed in the study were appropriate only for multielement design graphs. Thus, further investigation of methods for interpreting graphs from other designs is warranted.

**Inspection aids.** Previous efforts to improve visual inspection have traditionally focused on inspection aids such as regression and split-middle lines. These methods are used to estimate trends in the data, and the resulting lines
are drawn on the graph with the purpose of aiding analysis. Least-squares regression lines have the benefit of reliability, as they are calculated using the actual data values and precise equations, but they are difficult to compute by hand (Cooper et al., 2007). Much easier to draw are split-middle lines, but these represent an estimate of overall trend and are not precise (Cooper et al.). Additionally, the use of split-middle lines may result in unacceptably high Type I error rates (i.e., the risk of false positives; Fisher et al., 2003). Furthermore, the use of techniques that emphasize trend may influence the importance raters place on trend, to the exclusion or minimization of other relevant factors such as level and variability (Skiba, Deno, Marston, & Casey, 1989). Therefore, techniques that consider multiple factors that are important to visual data analysis are needed.

Normand and Bailey (2006) evaluated the accuracy of judgments of functional control on hypothetical treatment graphs with and without celeration lines. The authors presented AB- and ABA-design graphs to five Board Certified Behavior Analysts (BCBAs) and asked them to talk aloud as they inspected the data. Half of the 24 graphs included celeration lines, drawn on the graphs such that half of the data points in a phase fell above and half fell below the line. Graphs with and without celeration lines were presented in a multielement design. The BCBAs judged whether an increase or decrease in the target behavior was demonstrated in each graph, or noted that no change could be credited to the intervention. Additionally, they rated their confidence in these
decisions on a 7-point scale. The accuracy of the judgments was determined by their concordance with the predetermined answers based on the programmed graph characteristics (i.e., slope, mean shift, trend). Overall, the decision accuracy of the BCBAs was 72%. Interestingly, accuracy was poorer when celeration lines were present than when they were absent, although participants made more comments (in total, and regarding trend specifically) for graphs that included celeration lines. Consistent with previous studies on the reliability of visual inspection, Normand and Bailey found low accuracy and low agreement among raters. Their study extends previous research by including ABA-design graphs, a true experimental design that allows for determination of functional control. However, they found that inclusion of the final A phase in these graphs did not improve decision accuracy.

In a series of studies, Fisher et al. (2003) developed and evaluated the dual-criteria (DC) and conservative dual-criteria (CDC) methods to aid the visual inspection of data presented in single-case AB graphs. With the DC method, two computer-generated lines are drawn in the second phase of the graph, representing (1) an extension of the regression line for the data in the first phase, and (2) the mean level of the data in the first phase. An effect is demonstrated when the number of data points falling above both lines or below both lines meets the criterion determined by the length of the phase. This requirement is calculated using a binomial equation, and individuals inspecting the graphs using the DC method are supplied with a table of these requirements.
for phases of different lengths. The CDC method is the same as the DC method, with the adjustment that the two lines are raised by 0.25 standard deviations (see Figure 1 for an example graph with CDC lines drawn in the second phase).

Using Monte Carlo computer simulations, Fisher et al. (2003) demonstrated that the DC and CDC methods resulted in fewer errors than the historically prevalent split-middle method. Furthermore, they found that these methods wielded higher statistical power than the general linear model and interrupted time series, two statistical methods tested in the study.

In Study 1, Fisher et al. (2003) showed that the DC and CDC methods they had developed had sufficient power to detect true treatment effects while controlling error rates in the computer simulation. Moreover, of all the tested methods, the CDC method fared best for data sets of all lengths and was least affected by autocorrelation. In Study 2, five bachelor’s-level staff members of a behavioral treatment program for severe problem behavior used the DC method to evaluate AB-design graphs following individual instruction (lasting 10-15 min) on how to use the method during visual inspection. The five staff members averaged 55.4% correct visual inspection decisions in baseline, and improved to an average of 93.5% correct following training in the DC method. Having shown that the DC method was effective in a small sample of raters who had been trained individually, the researchers designed a study to evaluate a training program for a large group of participants. In Study 3, training on the DC
Figure 1. Sample graph with conservative dual criterion (CDC) lines drawn in the second phase.
method was provided via computer projector to 87 workshop attendees at a behavior analysis conference. For Group A, performance increased from 72% to 96% correct following training. This effect was replicated with Group B, the performance of which increased from 70% to 93% correct with training. Notably, the training component lasted only 15 min, even in large groups, demonstrating the efficiency of teaching the DC method. In summary, Fisher et al. developed, validated, evaluated, and made efficient an effective visual inspection aid.

**Teaching visual inspection.** Stewart, Carr, Brandt, and McHenry (2007) assessed visual inspection accuracy before and after a traditional lecture similar to what students often experience in undergraduate research methods courses in which single-case methodology is covered. After the lecture, participants were required to pass a concept quiz before continuing to the performance assessment where they visually inspected data depicted in graphs. Students were able to pass the quiz, but knowledge of the concepts did not translate to improved visual inspection performance. When the traditional lecture did not result in improvement of visual inspection skills for any of the six participants, the researchers taught them to use the CDC method described by Fisher et al. (2003). The CDC method resulted in universal improvements in accuracy, leading to performance above 90% correct in the CDC phase for each participant. However, when the CDC lines were removed from the graphs in a return-to-baseline condition, the gains in accuracy were lost. Finally, for participants who experienced a second CDC condition, performance accuracy improved rapidly
with the reintroduction of the CDC lines. The authors clearly demonstrated that
the CDC method was effective when it was used, but that the improvements
were lost when the aids were removed.

Visual inspection aids that modify graphs (e.g., the split-middle and CDC
methods), although effective, have disadvantages. Most prohibitively, they are
not portable. Thus, when inspecting data in journal articles or at professional
conferences, these aids are not available. Many involve complex equations that
are not easily memorized or are difficult to perform without computers (e.g.,
linear regression; Cooper et al., 2007). Furthermore, the improvements gained
using these techniques are lost when the aids are removed from the graphs.
Additionally, teaching just the concepts involved in visual analysis has been
shown to be insufficient for developing the appropriate repertoire (Stewart et al.,
2007). The studies above incited the development of a portable tool to teach the
process of visually inspecting graphed data for the current investigation.

The purposes of the present experiments were to assess the utility of a
portable tool for visual inspection (i.e., the job-aid; Experiment 1), and then to
determine whether the job-aid could be used effectively in a classroom
environment and, at the same time, compare it to a traditional lecture
(Experiment 2). The goals of Experiment 1 were (a) to observe and compare
student performance on visual inspection tasks with and without the job-aid
using a single-case research design, (b) to conduct a fading evaluation to explore
the maintenance of performance gains when the job-aid was removed, and (c)
to analyze participant errors and provide remedial training if indicated. The goals of Experiment 2 were (a) to compare the performance of students receiving traditional lecture to that of those taught to use the job-aid using a between-groups design, (b) to demonstrate the acquisition of conceptual knowledge via a concept quiz, (c) to replicate the findings by presenting the alternate training method to participants not meeting criterion performance following the first intervention, therefore producing a second set of data, and (d) to provide and evaluate the effectiveness of immediate feedback for participants not performing to criterion following exposure to both instructional methods.

EXPERIMENT 1: DEVELOPMENT AND VALIDATION OF THE JOB-AID METHOD

Participants

Seven undergraduate students from a large public university in the Midwest participated in the study. Their average age was 19.7 years (range, 18-24), and their average year in school was mid-freshman (range, freshman - 5+). Two participants (28.6%) were female. The average reported GPA (n=3) was 3.36 on a 4-point scale (SD = .276). Of the 6 participants who reported an academic major, 1 (16.7%) had declared psychology as a major, 2 (33.3%) were Aviation Flight Science majors, and 3 (50%) reported other majors as their academic foci (i.e., Nursing, French & Political Science, Business). All were enrolled in introductory classes in the psychology department of a state university with a total enrollment of approximately 25,000 students. All
participants reported that they had not taken a research methods course or worked as a research assistant where they were exposed to single-case design data. All participants provided informed consent prior to participation.

**Setting**

Sessions were conducted individually with each participant and occurred in small research/treatment rooms within the psychology department. These rooms were furnished with a table, chairs, lamp, and a picture on the wall for decoration. The rooms were equipped with video cameras for recording sessions and for supervision, but only the pilot participant’s sessions were taped. Sessions lasted 45-90 min, depending on the pace at which participants inspected the graphs. Participants attended 3-13 sessions; the number of sessions depended upon their performance (e.g., if participants reached criterion performance following written instructions plus teaching, they did not need to return for additional training or feedback) and whether they experienced job-aid fading.

**Materials**

*Graphs.* Five hundred different AB-design graphs depicting hypothetical data were created using Microsoft Excel® and the Resampling Stats® plug-in (see Figure 2 for an example). Graphs differed in level, trend, variability, and effect size, with half of the graphs demonstrating an effect and half demonstrating no true effect. When the graphs were created for the experiment, it was determined whether each graph depicted a true effect. First, effect sizes were programmed into the computer program (0 for non-effect graphs; 0.25 – 1.0 for effect
graphs), and secondly, the CDC method (Fisher et al., 2003) was then applied to each graph and indicated an effect or non-effect. Graph packets contained 10 graphs, with each packet containing 5 graphs that demonstrated an effect and 5 that did not. The effect and non-effect graphs were intermixed and their order randomized within each packet. Although each packet contained different graphs, the composition of the graph packets was consistent in the types of graphs included. For example, each packet contained a graph depicting high variability but no true effect, and each contained a graph with true differences in level and trend. The 10 types of graphs that comprised each packet are depicted in Table 1. To assess the accuracy of the graphs, experienced visual inspectors (i.e., behavior analysis faculty and graduate students) evaluated the stimuli used in the experiment to determine whether they depicted the qualities (i.e., differences in level, trend, and variability) they were designed to exhibit. The average agreement for each of the 10 graph types was 91.4% (range, 75-100%).

**Job-aid.** Written instructions were developed based on a task analysis of the process of visually inspecting AB-design graphs. The task analysis was derived from common text recommendations regarding visual inspection, rather than by observing or asking expert visual inspectors. The written instructions were compiled into a two-page job-aid with written prompts to evaluate level, trend, and variability in both baseline and treatment phases, and to compare
Figure 2. Sample graph presented to participants. This graph demonstrates a true effect with a change in level and no changes in trend or variability.
these dimensions across phases. Graphs were added to illustrate how to complete the steps (e.g., examples of trend estimation) and demonstrating exceptions to the guidelines described (e.g., graphs displaying changes in level but having ascending trends in both baseline and treatment, thus displaying no true effect of the independent variable). Lastly, a step was included to summarize the findings and make a decision regarding whether an effect was demonstrated in the graph under inspection. Following its development, the original job-aid was shown to two inexperienced undergraduate research assistants; their feedback was used to adapt the job-aid such that the instructions would be easier to follow by individuals unfamiliar with some technical terminology. The job-aid appears in Appendix A.

**Computer teaching presentation.** A computer-based training presentation was developed to enhance an instructional session with the researcher (see Appendix B). The presentation demonstrated use of the job-aid with several sample graphs and guided the viewers through the visual inspection process for each graph.

**Procedure**

In all conditions, participants indicated whether an effect was demonstrated in graphs created for the study by checking “yes” or “no” in the space dedicated for their answer at the bottom of each graph page.

**Baseline.** Participants completed 3 to 11 graph packets (10 graphs per packet) prior to any experimental manipulation. The number of packets per
Table 1.

Summary of statistics for the 10 types of graphs used in the experiment.

<table>
<thead>
<tr>
<th>Graph Code</th>
<th>Status</th>
<th>Description</th>
<th>Effect Size</th>
<th>Baseline Mean</th>
<th>Treatmen t Mean</th>
<th>Baseline Slope</th>
<th>Treatmen t Slope</th>
<th>Baseline Slope minus Treatment Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No Effect</td>
<td>ascending trends in both baseline and treatment</td>
<td>0</td>
<td>M = 54.59 SD = 7.33</td>
<td>M = 77.04 SD = 8.92</td>
<td>M = 0.07 SD = 0.03</td>
<td>M = 0.07 SD = 0.03</td>
<td>M = 0.00 SD = 0.00</td>
</tr>
<tr>
<td>2</td>
<td>No Effect</td>
<td>flat slope; no convincing change in level; little variability</td>
<td>0</td>
<td>M = 50.31 SD = 2.03</td>
<td>M = 50.45 SD = 1.93</td>
<td>M = 0.00 SD = 0.13</td>
<td>M = 0.01 SD = 0.15</td>
<td>M = -0.01 SD = 0.13</td>
</tr>
<tr>
<td>3</td>
<td>No Effect</td>
<td>flat slope; no convincing change in level; high variability</td>
<td>0</td>
<td>M = 57.24 SD = 8.74</td>
<td>M = 57.54 SD = 7.60</td>
<td>M = 0.00 SD = 0.03</td>
<td>M = 0.01 SD = 0.03</td>
<td>M = -0.01 SD = 0.03</td>
</tr>
<tr>
<td>4</td>
<td>No Effect</td>
<td>flat slope; no convincing change in level; moderate variability</td>
<td>0</td>
<td>M = 50.16 SD = 5.37</td>
<td>M = 50.16 SD = 5.10</td>
<td>M = 0.01 SD = 0.05</td>
<td>M = 0.01 SD = 0.04</td>
<td>M = 0.00 SD = 0.07</td>
</tr>
<tr>
<td>5</td>
<td>No Effect</td>
<td>descending trends in both baseline and treatment</td>
<td>0</td>
<td>M = 47.80 SD = 6.78</td>
<td>M = 29.60 SD = 5.01</td>
<td>M = -0.08 SD = 0.03</td>
<td>M = -0.08 SD = 0.03</td>
<td>M = 0.03 SD = 0.13</td>
</tr>
<tr>
<td>6</td>
<td>Effect</td>
<td>changes in direction of trend (e.g., flat slope to ascending) and level</td>
<td>0.25</td>
<td>M = 49.83 SD = 2.43</td>
<td>M = 59.43 SD = 23.71</td>
<td>M = 0.02 SD = 0.05</td>
<td>M = -0.04 SD = 0.27</td>
<td>M = 0.07 SD = 0.28</td>
</tr>
<tr>
<td>7</td>
<td>Effect</td>
<td>change in steepness of slope (e.g., descending with steeper slope to descending with flatter slope)</td>
<td>1.0</td>
<td>M = 55.79 SD = 8.79</td>
<td>M = 119.82 SD = 6.42</td>
<td>M = 0.05 SD = 0.02</td>
<td>M = 0.11 SD = 0.02</td>
<td>M = -0.06 SD = 0.02</td>
</tr>
<tr>
<td>8</td>
<td>Effect</td>
<td>change in variability</td>
<td>0*</td>
<td>M = 50.88 SD = 4.66</td>
<td>M = 51.04 SD = 3.85</td>
<td>M = -0.01 SD = 0.13</td>
<td>M = -0.02 SD = 0.08</td>
<td>M = 0.02 SD = 0.13</td>
</tr>
<tr>
<td>9</td>
<td>Effect</td>
<td>moderate change in level between baseline and treatment phases</td>
<td>1.0</td>
<td>M = 50.10 SD = 6.37</td>
<td>M = 78.43 SD = 7.02</td>
<td>M = -0.02 SD = 0.04</td>
<td>M = -0.02 SD = 0.04</td>
<td>M = 0.00 SD = 0.02</td>
</tr>
<tr>
<td>10</td>
<td>Effect</td>
<td>changes in level, steepness of slope, and variability</td>
<td>1.0</td>
<td>M = 49.18 SD = 7.61</td>
<td>M = 81.41 SD = 7.81</td>
<td>M = 0.02 SD = 0.02</td>
<td>M = 0.27 SD = 0.06</td>
<td>M = -0.25 SD = 0.06</td>
</tr>
</tbody>
</table>
participant depended on each participant’s performance; baseline data were collected until visual analysis of the data suggested steady-state behavior. Participants did not receive feedback on their performance in baseline.

**Job-aid only.** Following stable responding in baseline, participants were given the job-aid (one for each graph) and told to use it to help them evaluate the graphs. They were not provided with instruction on how to use the job-aid. Participants did not receive feedback on their performance in the job-aid only phase.

**Job-aid plus teaching.** Participants received the job-aid and experienced an experimenter-led session enhanced by the computer presentation in which they followed along with the models, wrote on job-aid worksheets and drew on sample graphs provided to them as the researcher delivered instructions and answered their questions. The presentation was used with live instruction as the researcher prompted participants to estimate mean and trend lines and to rate the variability of the data in baseline and treatment phases along with the model on the computer screen. The researcher, assisted by the visual aid of the computer presentation, led the participants through the job-aid for each sample graph, provided information about how to apply the written guidelines, and answered questions from the participants. Participants did not receive feedback from the researcher on their performance during this phase.

**Job-aid fading.** Two participants (Cindy, Emily) experienced job-aid fading after they had reached criterion performance with the full job-aid and
teaching. Fading the job-aid consisted of three steps: (1) the job-aid was decreased to a one-page worksheet (from two pages) in which the written instructions were abbreviated and the sample graphs were removed, (2) the job-aid was reduced to a half-page worksheet containing only the five main questions from the original two-page job-aid (e.g., “Did level change from baseline to treatment?”), and (3) the job-aid was removed entirely (however, no participants reached this step). The faded job-aids appear in Appendix C.

**Error correction.** An error analysis revealed that one participant (Aaron) who had not reached criterion performance with the job-aid plus teaching package was not consistently following the written instructions accurately. For example, he might write that the level of behavior in the baseline phase was 5 and that the level in treatment was 5, but then write that level was different between baseline and treatment. Error correction consisted of the researcher pointing out errors immediately when the participant wrote something incorrect or inconsistent with his previous answers on the same job-aid. The participant then corrected the error and continued completing the job-aid to come to a conclusion about whether an effect was demonstrated in the graph. The participant did not receive feedback about the accuracy of his judgments of treatment effects or non-effects during this phase.

**Modified job-aid.** The participant who received error correction (Aaron) also received a modified job-aid, which then became the final job-aid for participants in Experiment 2. The change consisted of modifying the five main
questions of the job-aid such that they read, “Is there a convincing change in ___ (e.g., variability) from baseline to treatment?” rather than asking for a rating of degree of change (no change, a little, a lot). This modification was made because the researchers observed that the participant would estimate graph characteristics (e.g., level) for both phases at very similar values but then answer that there was a change in that characteristic between phases, therefore leading him to an erroneous conclusion and increasing the percentage of false positives in his answers.

Trend training. Error analyses were conducted for the two participants whose performance did not reach criterion following the job-aid plus teaching package (Aaron, Rylan). The researchers noted that both participants often incorrectly estimated the slope or direction of trends (or lack of trends) in the data, and that this led them to erroneous conclusions about treatment effects. Thus, the experimenter conducted trend training individually with these participants to improve their trend-estimating skills. Participants were given one graph at a time and asked to draw a line estimating the trend. The experimenter then provided feedback on the accuracy of the line drawn by the participant. Trend training continued until each participant correctly estimated trend for five consecutive graphs. Rylan completed trend training in 29 graphs; Aaron completed trend training in 71 graphs.

Feedback. Despite improving Aaron and Rylan’s trend-estimating skills, their performance on the visual inspection task did not increase to criterion.
Therefore, these participants were given immediate feedback (including praise and corrective feedback) as they evaluated additional graphs. For example, based on a participant’s responses to a certain graph, the experimenter might have said, “You did a great job estimating trend! Both phases have ascending trends. However, you reported that the graph demonstrated an effect, and that is not the case because this graph depicts a ‘Level Exception’ as shown on the back of the job-aid. See, level was the only dimension that changed across phases, and if you cover the phase line, it looks like both paths could be part of the same data path.” Feedback continued for Rylan until he scored 90% on three consecutive graph packets. Feedback was terminated early for Aaron because he left the study before we could complete the evaluation.

**Results**

Visual inspection of graphically displayed data was the primary method of data analysis for Experiment 1. Graphed data for the seven participants appear in Figures 3-5. The three panels of Figure 3 display similar results. Depicted in the top panel, Eric’s data show a mean score in baseline of 50%. When the job-aid alone was introduced, his performance did not improve substantially ($M = 53.3\%$). However, when teaching with the computer presentation was added to the job-aid, his performance showed immediate substantial improvement ($M = 80\%$). In the middle panel, Dusty’s baseline data show a mean of 50%. His accuracy improved to 70% immediately following the introduction of the job-aid. Dusty’s results do not show a substantial further increase in performance when
teaching was added to the job-aid ($M = 72.5\%$). Josh’s data appear in the bottom panel of Figure 3. Across all baseline sessions, Josh’s accuracy was 58.2% on average (range, 40-70%). With the introduction of the job-aid alone, his average score increased to 73.3%. When teaching with the computer presentation was added to the job-aid, Josh’s performance improved further, to an average of 81.3%.

A similar pattern of results is evident in Figure 4, in that participants’ visual inspection performance improved with the job-aid alone or following training using a computer presentation on how to use the job-aid. Additionally, these participants experienced job-aid fading to assess maintenance of visual inspection skills in the absence of the job-aid. Cindy’s data are presented in the top panel of Figure 4. Cindy’s performance improved from 60% to an average of 75% when the job-aid alone was introduced, and increased again to 88.3% after she was taught to use the job-aid with the help of the computer presentation. Next, we began fading the job-aid with Cindy. During fading step 1, in which the job-aid was reduced from two pages to one page and no longer contained sample graphs, Cindy’s accuracy remained high at 88.8%. Thus, we progressed to fading step 2, in which the job-aid was further reduced to one-half page containing only the main questions. During this phase, Cindy’s accuracy declined immediately to an average of 76.7%. Although this percentage was still acceptable, it was a clear deterioration compared with her previous performance. Therefore, we stopped fading and returned to the full job-aid. In this condition,
Figure 3. The percentage of accurate visual inspection in baseline, job-aid only, and job-aid plus teaching conditions for Eric (top panel), Dusty (middle panel), and Josh (bottom panel).
Figure 4. The percentage of accurate visual inspection in baseline, job-aid only, and job-aid plus teaching conditions for Cindy (top panel) and Emily (bottom panel).
Figure 5. The percentage of accurate visual inspection in baseline, job-aid only, job-aid plus teaching conditions, and additional manipulations for Aaron (top panel), and Rylan (bottom panel).
Cindy’s accuracy increased to an average of 80%. The researchers would have preferred to continue working with Cindy to see if her previous level of accuracy could be recaptured, but her behavior in later sessions suggested she might not put forth her best effort in future meetings (she showed signs of boredom and frustration such as sighing, rocking, and turning pages flippantly), so her participation was discontinued after 13 sessions.

Emily’s data appear in the bottom panel of Figure 4. Like Cindy, she experienced job-aid fading following successful performance with the job-aid plus teaching package. In baseline, Emily’s mean accuracy was 56.7%, and her performance did not improve when given the job-aid alone (M = 52.5%). Following training to use the job-aid with the assistance of the computer presentation, Emily’s accuracy increased to an average of 80.9%, with 3 of the last 4 data points in that phase at 90%. We then began fading the job-aid. During fading step 1, Emily’s performance averaged 78%. With Emily’s accuracy moving in an undesirable direction, we returned to the full job-aid, where previous levels of success (M = 83.3%) were again observed.

Figure 5 depicts the data for the final participants, Aaron (top panel) and Rylan (bottom panel). Unlike the five previously discussed participants, the performance of these two participants did not improve with the introduction of the job-aid or with job-aid plus teaching with the computer presentation. In baseline, Aaron’s scores averaged 60% (range, 40-90%). When he was given the job-aid without any training on its use, his average was 57.1%, and when
training with the computer presentation was included, his accuracy was in the same range, with an average of 60%. Next, Aaron experienced error correction, in which the researcher corrected clerical errors he had been making on the job-aids that sometimes resulted in incorrect answers. Although Aaron’s clerical errors declined, his visual inspection performance did not improve ($M = 56.7\%$). Aaron was then given a modified job-aid (described above), but this did not have an effect on his performance ($M = 56.7\%$). An error analysis revealed that Aaron was often misjudging trends (and lack of trends) in the graphs under evaluation. Therefore, we conducted trend training with Aaron to remediate this deficit. Aaron was trained to conduct trend estimation with a high degree of accuracy, but still this did not result in a substantial improvement in visual inspection performance ($M = 61.4\%$). Finally, we began feedback, during which, immediately after Aaron made his decision, the researcher informed him whether his answer was correct and why. After three sessions of feedback, Aaron’s accuracy was 66.7%. Aaron left the study before this evaluation could be completed.

Rylan’s data are presented in the bottom panel of Figure 5. Like Aaron, Rylan’s performance did not improve substantially with the job-aid, with or without teaching with the computer presentation. Rylan’s accuracy averaged 72% in baseline, notably higher than all other participants in the study. With the job-aid alone, his performance was 65.7%, and after he was taught to use the job-aid with the computer presentation, accuracy was 76.7%. An error analysis
revealed that Rylan, like Aaron, was often estimating trend incorrectly, and that this led to many of his erroneous decisions of treatment effect/non-effect. Thus, we conducted trend training with Rylan and, although his trend-estimation skills improved, visual inspection performance did not ($M = 63.3\%$). Finally, we provided immediate feedback to Rylan as he evaluated graphs. With feedback, Rylan’s accuracy improved to an average of 84%.

**Discussion**

The job-aid, either alone or in combination with the lecture with computer demonstration, resulted in improved visual inspection accuracy over baseline performance in 5 of the 7 participants. Impressively, these participants improved from near-chance levels of accuracy to levels considered good by experts.

For the two participants whose performance was not enhanced by the job-aid and computer presentation, error analyses were conducted to determine what errors might be leading to incorrect decisions. These analyses revealed consistent mistakes and specific skills deficits (i.e., incorrectly completing the job-aid, poor trend estimation) that the researchers then remediated with supplemental training. However, despite improvement of those skills, visual inspection performance did not improve for one participant until immediate feedback was provided on his judgment of treatment effects. The remaining participant left the study before the end of the evaluation.
Two participants (Cindy, Emily) experienced the fading evaluation in which we began removing the job-aid to measure maintenance of performance improvement in its absence. Unfortunately, we observed decreases in accuracy during fading before removing the entire job-aid and returned immediately to the full job-aid. However, these decreases were small for both participants and it is unknown whether their performance would have continued to decline or would have stabilized at acceptable levels. With limited experience using the job-aid to inspect single-case design data, performance improvement did not maintain fully when we started fading the portable inspection tool. It is possible that fading may have been more successful if participants had been given more experience with the job-aid and with inspecting graphs prior to attempts to fade the job-aid. This might be an important direction for future research because if, with more experience, performance maintains when the job-aid is removed, it would have significant implications for the manner in which we teach these critical skills.

Despite the widespread perception that visual inspection expertise can be gained only through extensive experience, participants in Experiment 1 reached mastery levels in as little as one training session with the job-aid. The traditional way we learn to visually inspect graphically displayed data most often includes many instances of trial-and-error and is sometimes unpleasant (remember that time you mistakenly identified an effect in a graph and were corrected in front of the entire lab?). Instead, it may be easier and more efficient to have decision rules and tools to aid visual analysis.
In conclusion, the portable job-aid, either alone or following computer demonstration, improved visual inspection skills for most participants to levels similar to those obtained by groups of experts. The results demonstrated that individualized supplemental procedures may be used to remediate deficiencies in foundational skills involved in visual analysis (e.g., trend estimation), but that correcting those deficiencies may not result in better visual inspection accuracy. Instead, it may be more economical, and in accordance with traditional classroom practices, to simply provide corrective feedback to students performing at unacceptable levels.

EXPERIMENT 2: EVALUATION OF THE JOB-AID WITH TRAINING IN GROUPS

Method

Participants

Forty undergraduate students from a small private college in the northeastern United States participated in Experiment 2. Their average age was 19.0 years (range, 18-31). Twenty-two participants (55%) were female. Twenty-nine (72.5%) were freshman, eight (20%) were sophomores, two (5%) were juniors, none were seniors, and one (2.5%) listed her year in school as 5+. Nine students (22.5%) listed psychology as a major and one student (2.5%) listed psychology as a minor. The average reported GPA (n=37) was 3.18 on a 4-point scale ($SD = 0.458$), and participants were taking an average of 14.9 credits (range, 3-18) during the semester in which they participated. Participants reported working an average of 7.96 hours per week ($SD = 8.2$) outside of
school. Worth mentioning is that 23 (57.5%) of the participants reported that they were planning to apply to graduate school. Eight participants (20%) did not receive extra credit or another incentive for participation. All participants were enrolled in classes in the psychology department of a small private college with a total enrollment of approximately 3,700 students. Participants reported that they had not taken a research methods course or worked as a research assistant where they were exposed to single-case design data. All participants provided written consent to be included in the study prior to participation.

Potential participants who scored greater than 90% on the Pretest were excluded because they had demonstrated that they did not need to improve their visual inspection skills. Four participants were excluded for this reason. Their mean age was 18.25 years, and all were freshmen. One participant (25%) was female. Two (50%) had declared psychology as a major; one (25%) was undecided, and the remaining participant was majoring in secondary education and mathematics. Three (75%) reported that they planned to attend graduate school, and the other listed that he was unsure about attending graduate school. Excluded participants were taking a mean of 15.25 credit hours ($SD = 2.22$) that semester, and they reported working an average of 18.75 hours per week outside of school ($SD = 14.10$). The mean GPA of the 3 excluded participants (75%) who reported this information was 3.23 on a 4-point scale ($SD = .40$).
Setting

Sessions were conducted in typically furnished classrooms and conference rooms on campus. All sessions but one had 2 to 7 participants at a time. One session was conducted with only one participant because the participant had missed his group session and all other participants in the study had completed data collection. Sessions of traditional lecture lasted approximately 60 min, and sessions of job-aid plus teaching lasted approximately 90 min. Feedback sessions lasted from 45 to 120 min.

Power analysis

A power analysis was conducted using the average effect size generated by calculating Cohen’s d for each participant who received the job-aid plus teaching in Experiment 1. The power analysis indicated that a sample size of 13 participants per group would yield sufficient power (.8) to detect an effect of the independent variable consistent with effect sizes obtained from the previous study with an alpha level of .05. Thus, group sizes of 20 participants were used in Experiment 2 to ensure obtaining sufficient statistical power.

Experimental design

An alternative-treatments design with pretest (Shadish, Cook, & Campbell, 2002) between-groups design was used to compare the visual inspection performance of participants who experienced a traditional lecture with the performance of participants who received the job-aid and computer teaching package. Prior to participation, participants were randomly assigned to 1 of 2
teaching conditions. Participants not scoring at least 80% correct following one teaching condition experienced the other condition. Participants still not scoring 80% or higher after Posttest 2 received immediate feedback and evaluated additional graphs.

**Materials**

*Graphs*. The 500 graphs that were created for Experiment 1 were also used in Experiment 2.

*Traditional lecture*. A videotaped lecture was developed for the study in which an instructor presented material related to the visual inspection of single-case design graphs, including descriptions of the concepts of level, trend, and variability, and covering the subject matter suggested by one of the most popular textbooks in applied behavior analysis (Cooper et al., 2007). The lecture lasted approximately 20 min (see Appendix D for an outline) and included a PowerPoint presentation with key points and stylized graphs depicting variations in level, variability, and the direction and slope of trend within and across phases (see Appendix E). Prior to its use in experimental sessions, procedural fidelity of the videotaped lecture was assessed by two independent raters using a checklist designed to determine the extent to which the video covered what it was intended to cover (see Appendix F). Procedural fidelity for the video was 100%.

*Job-aid*. The job-aid for the visual inspection of AB-design graphs developed in Experiment 1 was used in the present experiment (see Appendix A).
Computer teaching presentation. The computer teaching presentation developed for Experiment 1 was used in the present experiment (see Appendix B).

Concept quiz. A brief, 7-question multiple-choice quiz was developed to assess participants’ knowledge of key concepts following exposure to the teaching conditions (see Appendix G).

Dependent measures

The percentage of correct judgments of treatment effects and non-effects of 20 AB-design graphs served as the primary dependent measure of skill acquisition. Participants in both groups completed two packets of graphs (with 10 graphs per packet) prior to being exposed to one of the teaching conditions. Following training, all participants completed two additional packets of graphs. The average performance of participants in each group was evaluated statistically (described later). The number of participants scoring 80% or higher in each group served as a secondary dependent variable. Knowledge of key concepts was assessed by a brief quiz. Accuracy on the quiz was measured as the percentage correct out of 7 questions.

Procedure

See Figure 6 for a flowchart depicting the experiment’s conditions. Prior to the Pretest, participants were randomly assigned to Group A or Group B using a computer-based random sequence generator. The experimenter assigned each participant a number and put those numbers into the computer program. The
experimenter then took the resulting list and assigned the top half to Group A and the bottom half to Group B.

**Pretest.** Participants were given two packets of graphs and asked to determine whether each graph demonstrates an effect. Participants did not receive feedback regarding their performance. Participants scoring greater than 90% on the Pretest were excluded from the study because they demonstrated that they did not need further instruction in visual inspection.

**Traditional lecture (Group A).** Participants in this condition viewed the videotaped lecture on visual inspection. The researcher administering the video then answered questions from participants as might be done in a college lecture, but did not provide individual instruction on graph analysis.

**Job-aid plus teaching (Group B).** Participants received the job-aid and experienced an experimenter-led session enhanced by the computer presentation in which they followed along with the models, wrote on job-aid worksheets and drew on sample graphs provided to them as the researcher delivered instructions and answered their questions. The presentation was used with live instruction as the researcher prompted participants to estimate mean and trend lines and rate the variability of the data in baseline and treatment phases along with the model on the computer screen. The researcher, assisted by the visual aid of the computer presentation, led the participants through the job-aid for each sample graph, provided information about how to apply the written guidelines, and answered questions from the participants.
**Concept quiz (both groups).** All participants completed a brief quiz on the concepts involved in visual inspection following exposure to their first treatment condition (see Appendix G).

**Posttest 1 (both groups).** Following exposure to traditional lecture or the job-aid plus teaching, all participants evaluated two packets of graphs. Participants from the condition including the job-aid were asked to use their job-aids to help them evaluate the graphs. Participants did not receive feedback regarding their performance. Group performance was then compared. Individual performance data provided a supplementary measure of intervention effect, showing the performance of individual participants and permitting further investigation of unexpected or anomalous findings. Individual performance data were also used to identify students to receive additional practice with immediate corrective feedback following inadequate performance (i.e., less than 80% accuracy) on the Posttests and was used to assess the effects of this feedback on visual inspection performance.

Participants who did not score 80% or above on Posttest 1 received the intervention they had not previously experienced and then took Posttest 2. Participants still not scoring 80% or above then received feedback in individual sessions.

**Feedback (both groups).** Participants whose accuracy on the Posttest graphs following exposure to both teaching conditions was less than 80% received individual performance feedback, including praise and error correction,
Figure 6. Flowchart depicting the sequence of study procedures.
as they evaluated additional graphs as described in Experiment 1. Feedback continued until participants reached 80% accuracy or completed six graph packets. Participants had access to the job-aid as they evaluated the graphs.

Results

**Interrater agreement**

One hundred percent of the graph packets across all tests (i.e., Pretest, Posttest 1, Posttest 2, Feedback) were scored by an independent rater. An agreement was defined as both raters recording the same score for a graph packet. Point-by-point agreement was used to determine the interrater agreement score and was calculated by dividing the number of agreements by the number of agreements plus disagreements and then multiplying by 100. Interobserver agreement was 99.2% (range, 90 – 100%).

**Concept quiz**

Participants in Group A scored a mean of 4.9 out of 7 points (70%) on Quiz 1 ($SD = 1.25$) following the traditional lecture. Participants in Group B scored a mean of 5.05 (72.14%) following the job-aid plus teaching condition ($SD = 1.23$). These results were not significantly different using an independent samples $t$-test, $t(38) = -0.381, p = .705$. These data suggest that participants learned the concepts of visual inspection to the same degree regardless of whether they experienced the traditional lecture or the job-aid plus teaching condition. On Quiz 2, participants remaining in Group A scored a mean of 4.67 ($SD = 1.72$) following the job-aid plus teaching package. The remaining Group B
participants scored a mean of 5.3 points ($SD = .95$) on Quiz 2 after watching the traditional lecture video. These results were not significantly different, $t(20) = -1.036$, $p = .313$, nor do they indicate improved conceptual repertoires following the additional intervention.

Data analysis

The experimental results were evaluated through statistical analysis of the between-groups data. Additionally, the data were graphed such that changes in means and standard deviations were evident (see Figures 7 and 8). The percentage of participants scoring 80% or higher on each test are depicted in Figure 9.

**Statistical analysis.** An independent samples $t$-test was performed on the Pretest data comparing Groups A and B to determine whether statistically significant differences existed between the groups prior to intervention. A related-samples $t$-test was performed on the data from Group B comparing Pretest and Posttest scores to identify a main effect of the job-aid plus teaching package. A related-samples $t$-test was performed on the data from Group A comparing Pretest and Posttest scores to identify a main effect of the traditional lecture. An independent samples $t$-test was performed on the Posttest data comparing Groups A and B to identify any differences in the groups that could be attributed to the different interventions. A chi-square test for independence was conducted on the number of participants in each group scoring 80% or higher on
Posttest 1 to identify a potential relation between group membership and the number of participants meeting criterion-level performance.

**Test assumptions.** The data met the assumptions for the independent samples t-tests in the following ways: (1) participants were randomly assigned to groups, (2) the dependent variable is assumed to have a normal distribution in the population, and (3) the variance of one of the populations involved is not likely to be more than twice that of the other (Cohen, 2001) and the sample sizes were equal by assignment, thus satisfying the assumption of homogeneity of variance. Having met the three assumptions of the independent samples t-test, the data also necessarily met the two major assumptions for the related-samples t-test (i.e., normality and independent random sampling; Cohen).

The numbers of participants in each group scoring 80% or higher and, conversely, scoring less than 80%, were subjected to a chi-square test. These categorical data of the number of participants in each group scoring 80% or higher at Posttest 1 met the assumptions for the chi-square test of independence in the following ways: (1) group membership was mutually exclusive and exhaustive, (2) the observations were independent, and (3) the expected frequency of each cell was greater than five (Cohen, 2001).

**Findings.** Participants’ scores on the Pretest, Posttest 1, Posttest 2, and Feedback Posttests appear in Figures 7 and 8. Group means were compared with independent and related samples t-tests using SPSS Statistics (“SPSS,” 2008). Groups A ($m_A = 61\%, SD = 10.59$) and B ($m_B = 64.75\%, SD = 11.29$)
Figure 7. Group means for participants in Group B (top panel) and Group A (bottom panel). Error bars represent one standard deviation from the mean. The number of participants remaining in each group at Posttest 2 appears in parentheses.
Figure 8. Data for individual participants who received feedback in Group B (top panel) and Group A (bottom panel). Within each panel, each marker style represents one participant throughout all phases.
Figure 9. The percentage of participants in each group scoring 80% or higher at each test. Participants scoring 80% or higher on any test were considered to have met criterion and were scored as meeting criterion on all subsequent tests. For Group A, Posttest 1 followed the traditional lecture. For Group B, Posttest 1 followed job-aid plus teaching. Posttest 2 followed the job-aid for Group A, and followed the traditional lecture for Group B. The number of participants represented by each bar appears in parentheses.
performed similarly on the Pretest, $t(38) = -1.083, p = .286$. Following the traditional lecture, participants in Group A improved their scores to an average of 72.75% ($SD = 13.02$). This was a statistically significant increase from the Pretest, indicating a main effect of the traditional lecture, $t(19) = -3.113, p = .006$. Following training with the job-aid plus teaching package, participants in Group B improved their visual inspection scores to an average of 74.25% ($SD = 13.60$). This was also a statistically significant increase from the Pretest, indicating a main effect of the job-aid plus teaching package, $t = -2.510, p = .021$. Posttest 1 scores for both groups were significantly higher than Pretest scores ($t(39) = -4.018, p < .0001$), but did not differ significantly from one another ($t(38) = -0.356, p = .724$). Participants who did not score 80% or higher on Posttest 1 went on to experience the alternate intervention and took Posttest 2. Following training with the job-aid plus teaching package (after having experienced the traditional lecture), participants remaining in Group A ($n=13$) scored an average of 76.15% accuracy on Posttest 2 ($SD = 11.21$). After viewing the traditional lecture (after receiving training with the job-aid plus teaching package), participants remaining in Group B ($n=11$) scored an average of 73.64% on Posttest 2 ($SD = 12.47$). As in the previous comparison, these scores were not significantly different, $t(22) = .521, p = .608$. Participants still not scoring 80% or higher following the second teaching condition received immediate feedback as they evaluated additional graphs. During the feedback phase, participants
remaining in Group A (n=5) averaged 66.47% correct judgments ($SD = 3.28$). Participants remaining in Group B (n=7) scored an average of 76.19% ($SD = 9.37$). Levene’s test for equality of variances was significant, $F = 5.00, p = .049$. Thus, the results of the t-test with equal variances not assumed were used. This difference was statistically significant, $t(7.889) = -2.538, p = .035$, meaning that participants in Group B scored higher on the Feedback Posttests than their counterparts from Group A.

The percentage of participants scoring 80% or higher on each test is depicted in Figure 9. For Group A, which experienced the traditional lecture first, the following percentages of participants scored 80% or higher on the Pretest, Posttest 1, Posttest 2, and Feedback Posttests: 10%, 35%, 75%, and 100%. Interestingly, the number of participants in Group A scoring at the criterion level more than doubled from the traditional lecture levels following training on the job-aid plus teaching package. For Group B, which experienced the job-aid plus teaching package first, 15%, 55%, 65%, and 95% of participants scored 80% or higher on the Pretest, Posttest 1, Posttest 2, and Feedback Posttests, respectively. Therefore, the point at which both groups had the largest increase in the number of participants scoring 80% or better was following the job-aid plus teaching condition, regardless of the order in which they experienced the instructional methods. However, chi-square analyses conducted on these data showed no significant relations between these variables. Eleven of the 40 participants (27.5%) required corrective feedback to reach criterion performance,
and one participant (2.5%) did not reach criterion performance within six feedback sessions.

Discussion

The purpose of Experiment 2 was to determine whether students taught to use written instructions (i.e., the job-aid) to visually inspect single-case design data would perform better on visual inspection tasks than students who received a traditional lecture covering the conceptual framework of visual inspection. Results indicated main effects for both treatments, but similar mean visual inspection accuracy between groups. Interestingly, further examination revealed that more than twice as many participants (16 vs. 7) reached criterion performance following the job-aid plus teaching package than following the traditional lecture. However, this finding was not considered significant when subjected to statistical testing.

The results differ from previous work (Stewart et al., 2007) in that in the current investigation, the performance of participants who experienced the traditional lecture improved from their earlier (Pretest) scores and did not differ from participants who used a visual-analysis aid. In contrast, Stewart et al. found that traditional lecture did not improve the visual inspection skills of any of their six participants, even though participants in that study demonstrated acquisition of the conceptual knowledge related to visual inspection. The difference in results in Experiment 2 and the Stewart et al. investigation cannot be accounted for by a ceiling effect because participants in both studies had comparable
baseline and pretest scores (approximating chance levels). One possible explanation for this discrepancy could be that the videotaped lecture used in the current study was superior in some way to the one used in the study by Stewart et al. However, both videotaped lectures were based on material from the foremost textbook in applied behavior analysis (i.e., Cooper, Heron, & Heward, 1987, 2007) and were similar in content. It is possible, though, that an addition to the lecture for the current study could have influenced the results. The lecture for the current study included a PowerPoint presentation with several slides detailing how to use the split-middle technique to estimate trends in the data. The lecture from the Stewart et al. investigation did not include PowerPoint slides, so participants in that study did not view similar slides providing step-by-step instruction on trend estimation. Three participants (15%) in Group A of the current experiment drew lines on their graphs, suggesting that participants may have been influenced by these slides. The lectures differed in other ways as well, which may have impacted participant responses. The videotaped lecture from Stewart et al. was evaluated using the fidelity checklist for the traditional lecture video developed for the current investigation (see Appendix F). As assessed with this measure, the video from Stewart et al. met 75% of the fidelity criteria, compared with 100% for the video in the current study. Additionally, the lecture from Stewart et al. lasted approximately 6 min, whereas the lecture from Experiment 2 of the current investigation lasted approximately 21 min. Given the large difference in duration, it is reasonable to conclude that more information
was provided by the lecture in the current study. However, without an experiment comparing these videotaped lectures directly, it will remain unclear why participants in the traditional lecture group of the current investigation scored higher on visual inspection performance tasks than did participants in the Stewart et al. study following exposure to a traditional lecture.

An alternative potential explanation for the difference in outcomes is the different populations studied in previous and current research. It is possible that students at a small private college like those who participated in Experiment 2 may be more likely to benefit from lecture than students at a large public university, such as participants in Stewart et al. (2007). Unfortunately, detailed demographic data (e.g., GPA) from the Stewart et al. investigation are unavailable. Future investigations should comprehensively describe participant characteristics to facilitate between-study comparisons.

A third possible cause for the differences in results between the current experiment and Stewart et al. (2007) is sampling error. Six students participated in the investigation by Stewart et al.; none responded to traditional lecture. Eleven of 31 participants (35.5%) who experienced the traditional lecture in the current study did not respond to the lecture. It is possible that sampling error could account for the results obtained in the investigation by Stewart et al., and that with larger sample sizes, different results might be obtained.

Experiment 2 results did not show differences in mean accuracy between participants who experienced the job-aid plus teaching and those who viewed
the traditional lecture. However, substantially more participants reached criterion-level performance (80% accuracy or better) following the job-aid plus teaching condition than following the traditional lecture. This measure is important because it mediates a principle concern with using group averages, specifically variability between participants. Scores of individual participants on each test varied widely. The scores of some participants improved greatly, whereas some scores did not change or even worsened following exposure to a teaching condition. Averaging these values resulted in similar mean accuracy percentages between groups, but the variability in scores within groups is evident in the high standard deviations. Therefore, the number of participants reaching criterion performance following each teaching method is informative. These data showed that the job-aid plus teaching package produced more criterion-level performance than the traditional lecture (i.e., 16 vs. 7 participants).

Additionally, participants in the job-aid group responded better to feedback later than participants in the traditional lecture group. The variable(s) responsible for this difference are unknown, but the difference indicates that the job-aid plus teaching package may have additional benefits to its recipients.

Statistically significant differences were not found between groups on Posttest 1. One possible explanation for the lack of difference between the mean performance of participants in the job-aid and lecture groups is the effects of the group setting on training. It is possible that because training was conducted in
groups, the job-aid instruction may not have been as responsive to participant behavior as is permitted in individual training sessions (as in Experiment 1). For example, the researcher may have been unable to monitor each participant’s attention and participation in the group setting. Some students may have ignored the lesson and worked ahead on their graphs and worksheets, potentially missing important instructions. In addition, some students may not have first written on their sample graphs during the third example in the computer presentation before the researcher showed the answer. Thus, we cannot know if potential reinforcement was occurring in the form of participants writing on the job-aids and then seeing their answers match the ones on the screen. Another possibility is that some students may not have completed their job-aids fully for each graph during the Posttests, potentially changing the decisions of treatment effects/non-effects and thereby altering their scores. In sum, the group setting could have provided many opportunities for all participants to not “receive” the same intervention. However, participants in the job-aid group in Experiment 2 performed similarly to participants who received the job-aid plus teaching package in Experiment 1; therefore, it is unlikely that the effects of the group setting alone could explain the absence of statistically significant differences between the job-aid and traditional lecture groups. It is important to note that, if these problems existed in Experiment 2, they were outweighed by the main effect of the job-aid. That is, the effects of the job-aid
were robust in that they persisted in even imperfect conditions - a desirable quality for all behavior-analytic interventions.

Indeed, potential benefits of the job-aid plus teaching package exist, even with finding results equivalent to those produced by the traditional lecture in Experiment 2. The computer teaching presentation took approximately the same time to administer as the traditional lecture, and did not require substantial preparation on the part of the educators because the presentation came packaged with the job-aid worksheets. Furthermore, training was brief, the job-aid was easy to use, it produced quick and educationally significant results, it was portable, and it was effective under less-than-ideal circumstances (i.e., in groups). Results suggest that professors may be able to teach the process of visual inspection using written instructions and a computer teaching model, and that students of behavior analysis may benefit from this departure from traditional lecture.

Experiment 2 contributes to the scientific knowledge regarding the teaching of visual inspection in that it showed that the job-aid could be used efficiently and successfully in the real-world setting in which it would most likely be used if adopted by teachers of behavior analysis. The experiment also demonstrated that feedback alone (without remediation of deficient foundational skills as in Experiment 1) could improve visual inspection accuracy in participants who did not respond to typical or enhanced treatment (i.e., lecture, job-aid package). Providing performance feedback is relatively easy and naturally occurs
in educational settings. A final strength of Experiment 2 is that participants who did not respond to the first intervention were then exposed to the alternate intervention, so that everyone who needed it received a chance to learn from both methods. Furthermore, if neither intervention sufficiently improved their performance, participants received individual feedback so they could benefit from their participation in the study.

Limitations of Experiment 2 must also be addressed. First, few measurements of behavior per condition were assessed for each participant. Repeated measurement is a hallmark of behavior-analytic research for reasons delineated in the introduction. Additionally, averages were used to compare group performance, which necessarily implies that some information is lost. This is particularly troubling given what may be high levels of variability across participants, as indicated by the large standard deviations for test scores. The reason for this variability is unknown because the data are collapsed into averages and this type of group design was not meant to explore within-group differences. It is also possible that much could be learned from individual participant data (e.g., the influence of some other, unknown, variable). Next, statistical analyses could mask some information, again because the data are abbreviated with statistical methods. Finally, skill maintenance was not assessed over time or in the absence of the job-aid; this may be a fruitful avenue for future research.
Like Hagopian et al. (1997), the present experiments demonstrated that written instructions can be used to improve the reliability of visual inspection. The current studies serve as an extension of this work from multielement designs to other single-case designs. Inexperienced students from Experiments 1 and 2, though from different populations and in different training settings, improved their visual inspection performance to levels comparable to experienced and expert judges. The effects of the job-aid were therefore replicated, providing convincing evidence that this portable tool can improve visual inspection skills in students for which these skills are deficient as well as for students inexperienced in analyzing single-case design data. Additionally, the job-aid is portable. This is a substantial benefit over other visual inspection aids, which have been shown to produce accurate results but that are impractical to use in certain crucial situations (e.g., when consuming published research in journals, when evaluating data in presentations at conferences).

Strengths of the current research are evident. An important and infrequently seen strength in this line of research is that we did not compare participant responses to those of experts. Rather, we programmed true effects and non-effects with a computer program and confirmed them with the CDC method (Fisher et al., 2003). Additionally, the current research was conducted in a practical setting, and application to the real world need not be inferred. The current research includes strong measurement methods, sufficient interobserver
agreement, robust procedural fidelity, and demonstrated integrity of the materials. We validated a method we had created to aid visual inspection and then increased the efficiency of its training and demonstrated its generality with different populations and in different settings.

Some limitations of the current research warrant attention. Many of the same criticisms that have been levied against other research on visual inspection apply. In the current studies, the task for participants was to determine whether change had occurred between phases. As in other studies that used AB-design graphs rather than true experimental designs, participants were unable to assess whether any observed change was attributable to the intervention. However, previous research by Normand and Bailey (2006) showed that including a final A phase (an experimental ABA withdrawal design) did not improve visual inspection accuracy. Furthermore, the comparison of data between two phases or conditions is necessary for functional control identification in all single-case designs. At the hub of additional criticisms from other research on visual inspection is that visual analysis in the studies is not conducted as it occurs naturally. In the real lives of clinicians and researchers, visual inspection is ongoing rather than occurring with completed graphs. However, one could successfully argue that in the midst of a research project is not the only time one might need to inspect data; for most people evaluating interventions, this occurs after data collection is complete (e.g., in articles and presentations). Finally, in real life, contextual information is available and influences visual analysis. The
current studies did not use real data, and no background information about the
treatment or behavior under study was provided. This is a valid concern, but one
I suspect will not be remediated until we are further along in this line of
research.

Future research should evaluate modified job-aids to guide visual
inspection for additional single-case designs. The current job-aid could be
modified to guide the user to assess within-phase changes and to include a
section on determining whether observed change between phases could be
attributable to the independent variable. Additionally, the current job-aid could
be strengthened by the addition of points that would prompt the user to consider
overlap of data between phases and to give less weight to outliers in the data.
For reversal designs, instructions could be added to address the pattern of
changes in responding in A phases compared to B phases, including whether the
target behavior changed during treatment phases and returned to baseline levels
in subsequent baseline phases. A job-aid for analyzing data in multiple baseline
designs would contain all the elements of the current job-aid plus instruction on
how to evaluate the replication of effects across panels. For example, a section
could be added to prompt the user to assess the extent to which behavior
changed when and only when the independent variable was manipulated.
Additionally, it could include questions to guide the inspector to observe the
consistency of the changes in responding across panels, including phenomena
such as delayed or temporary effects. For multielement designs, the individual
inspecting the data would compare graph characteristics between data paths rather than between phases, and the job-aid could be modified specifically to inform users how to respond to the separation of data paths, as well as how to compare test conditions to a control condition in the case of functional analysis data. For changing-criterion designs, the job-aid would be similar to the one guiding reversal designs in that the analyst would compare responding under the different levels of the independent variable (rather than between baseline and treatment phases). The individual inspecting the data would determine whether behavior reached the levels specified by each manipulation of the independent variable when and only when the criterion changed. As these examples illustrate, the job-aid could, in theory, be modified to address all single-case designs; future research should assess its generality to these different designs.

Visual inspection is the primary means of data analysis in our profession. The teaching of visual inspection to students of behavior analysis should be a focus of teachers in the field, and we should teach it using behavioral principles and strategies that have been proven effective. Visual inspection is complex, and some argue that the skills can be gained only through extensive experience. Notably, in the current studies, we have demonstrated high levels of accuracy in students with a fraction of that experience. The present studies are part of an emerging line of research that will eventually determine the utility, efficiency, and generality of tools to aid visual inspection, as well as assess the maintenance of skills following training or extended experience with such tools. Ultimately, this
line of research may have implications for how we as a field train future behavior analysts.
REFERENCES


Appendix A

Visual Inspection Job-Aid

Step 1: Level

Baseline:
- Draw a straight horizontal line with your eyes that leaves approximately half of the data points above it and half below. What y-axis number does this line cross? ______

Treatment:
- Draw a straight horizontal line with your eyes that leaves approximately half of the data points above it and half below. What y-axis number does this line cross? ______

* Compare the last few points of Baseline to the first few points of Treatment. Was there an immediate change in level between phases? Yes / No

☐ Is there a convincing difference between the levels (y-axis values) of the data paths in Treatment and Baseline? Yes / No

Step 2: Trend

Baseline:
- Draw a trend line with your eyes that represents the direction (up, down, flat) that leaves approximately half of the data points above it and half below.
  - What is the trend? ascending (up) descending (down) no trend (flat)

Treatment:
- Draw a trend line with your eyes that represents the direction (up, down, flat) that leaves approximately half of the data points above it and half below.
  - What is the trend? ascending (up) descending (down) no trend (flat)

* Is there a change in direction of trend from Baseline to Treatment? Yes / No
  * Is the trend steeper, flatter, or neither steeper nor flatter in Treatment compared to Baseline? steeper flatter neither

☐ Is there a convincing change in trend overall (direction or slope or both) between Treatment and Baseline? Yes / No

Step 3: Variability

**Baseline:** How far away from your imagined trend line are most of the data points?

Very near -- 1 2 3 -- Very far

**Treatment:** How far away from your imagined trend line are most of the data points?

Very near -- 1 2 3 -- Very far

![Variability Diagram]

☑ Is there a convincing difference in variability between Treatment & Baseline? Yes / No

Step 4: Make a Decision

**a. Summarize**

Was there a convincing change in level? Yes / No
Was there a convincing change in trend? Yes / No
Was there a convincing change in variability? Yes / No

*If you answered "No" for ALL of these, skip the Level Exception section and answer NO in (c) below.*

**b. Level Exception**

1. Was level the only dimension that changed? Yes / No
   If NO, SKIP the rest of the Level Exception section and answer YES to (c) below.
2. Did the trend stay the same in Baseline and Treatment? Yes / No
   If NO, SKIP the rest of the Level Exception section and answer YES to (c) below.
3. Was the trend ascending or descending (NOT flat) in both Baseline and Treatment? Yes / No
   If NO, SKIP the rest of the Level Exception section and answer YES to (c) below.
4. If you take away the phase line on the graph, does it look like the data points are part of the same data path? Yes / No
   If NO, answer YES to (c) below.

Examples:

![Examples Diagram]

*If you answered “YES” to Level Exception questions 1, 2, 3 and 4 above, answer NO in (c) below.*

**c. Did behavior change from Baseline to Treatment?** Yes / No

Transfer this answer to your graph packet.
Appendix B

Computer Teaching Presentation

Visual Inspection Job-Aid

Step 1: Level

Baseline: What is the mean level?
1. Draw a straight horizontal line with your eye that shows approximately half of the data points above it and half below.
What y-axis number does the line cross? 5

Step 1: Level

Treatment: What is the mean level?
1. Draw a straight horizontal line with your eye that shows approximately half of the data points above it and half below.
What y-axis number does the line cross? 5

The last few points in Baseline are at a different level than the first few points in Treatment.

Step 1: Level

Baseline: What is the mean level?
1. Draw a straight horizontal line with your eye that shows approximately half of the data points above it and half below.
What y-axis number does the line cross? 5

Step 1: Level

Treatment: What is the mean level?
1. Draw a straight horizontal line with your eye that shows approximately half of the data points above it and half below.
What y-axis number does the line cross? 5

The last few points in Baseline are at a different level than the first few points in Treatment.

Is there a consistent difference between the levels (y-axis values) of the data points in Treatment and Baseline? Yes No

- Baseline: y-axis value = 5
- Treatment: y-axis value = 5

No, they are the same.

Step 2: Trend

Baseline: What is the trend?
1. Draw a trend line with your eye that represents the direction (up, down, flat) that shows approximately half of the data points above it and half below.
2. What is the trend?
- ascending (up)
- descending (down)
- no trend (flat)

No trend (flat)

72
Step 2: Trend

**Treatment:** What is the trend?

1. Draw a trend line with your eye that represents the central tendency of your data. (up, down, flat) Don't worry about exactly where it is, just try to find the trend.

2. What is the trend?
   - **Ascending (up)**
   - **Descending (down)**
   - **No trend (flat)**

---

**Trend in treatment steeper**

**Trend in treatment flatter**

**Trend in treatment neither steeper nor flatter**

---

Step 2: Trend

Is the trend steeper, flatter, or neither steeper nor flatter in Treatment than in Baseline?

- Steeper
- Flatter
- Neither

---

The trend is steeper in Treatment.
Step 2: Trend
Is there a change in direction of trend from Baseline to Treatment? Yes/No
- Baseline: No trend (flat)
- Treatment: Ascending (up)
  Yes, they are different.
Is there a change in slope (steepness) of the trend in Treatment compared to Baseline? Yes/No
- Treatment Steeper
  Yes, there is a change in slope.

Step 3: Variability
Baseline: How far away from your imagined trend line are most of the data points?
- Very near - 1 2 3 - Very far

Step 3: Variability
Treatment: How far away from your imagined trend line are most of the data points?
- Very near - 1 2 3 - Very far
Step 3: Variability

Is there a convincing difference in variability between Treatment and Baseline? Yes (No)
- Baseline: variability rating = 1
- Treatment: variability rating = 1

No, they are the same.

Step 4: Make a Decision

Summarize
Was there a convincing change in trend? Yes (No)
- Baseline: No trend (flat)
- Treatment: Ascending (up); Steeper

Was there a convincing change in level? Yes (No)
- Baseline: y-axis value = 5
- Treatment: y-axis value = 5

Was there a convincing change in variability? Yes (No)
- Baseline: Very Near
- Treatment: Very Near

If you answered "No" for ALL of these, skip the Level Exception section and answer NO in (c) below

LEVEL EXCEPTION
If level was the only dimension that changed, AND the trend was the same in Baseline and Treatment (ascending or descending, NOT flat), it is a Level Exception graph.

The questions on the next slide and on your job-aid will help you determine if the level exception applies.

These are level exception graphs
**Step 4: Make a Decision**

If you removed the phase line, a Level Exception graph will look like the data paths could be part of the same line, like the graph on the right.

**Step 5: Make a Decision**

1. Was there a change in level? Yes \( \rightarrow \) No
2. If NO, did the rest of the Level Exception criteria hold? \( \rightarrow \) Yes \( \rightarrow \) No
3. Did the trend change or decreasing behavior hold in baseline and treatment? Yes \( \rightarrow \) No or \( \rightarrow \) Yes
4. For the level in the phase line on the graph, does the data points look like the data points are part of the same data path? Yes \( \rightarrow \) No

If YES in 1, 2, 3, and 4, answer NO in (c) below.

(c) Did behavior change from Baseline to Treatment?

**Our Example**

- Was there a convincing change in level? Yes (No)
- Was there a convincing change in trend? Yes (No)
- Was there a convincing change in variability? Yes (No)

**Let’s do another example...**

Here is the graph.

**Step 1: Level**

**Baseline** What is the mean level?

- Draw a straight horizontal line with your eyes that leaves approximately half of the data points above it and half below. What y-axis number does this line cross? 7

**Treatment** What is the mean level?

- Draw a straight horizontal line with your eyes that leaves approximately half of the data points above it and half below. What y-axis number does this line cross? 3
Step 1: Level

- Is there a convincing difference between the levels (y-axis values) of the data points in Treatment and Baseline? (Yes/No)
  - Baseline y-axis value = 7
  - Treatment y-axis value = 3
  - Yes, they are different

No, the level is approximately the same

Step 2: Trend

- Is the trend steeper, flatter, or neither steeper nor flatter in Treatment than in Baseline? (steep/
  - Baseline Descending (down)
  - Treatment Descending (down)
  - No, they are the same

- Is there a change in slope (steepness) of the trend in Treatment compared to Baseline? (Yes/No)
  - Treatment: Neither steeper nor flatter

Both trends are descending at approximately the same steepness
Step 2: Trend

- Is there a converging change in trend overall (direction or slope or both) between Baseline and Treatment? Yes (No)
  - Direction: No, they are both descending.
  - Slope: No, they are descending at about the same steepness.

Step 3: Variability

Baseline: How far away from your imagined trend line are most of the data points?
- Very near - 1 2 3 - Very far

Treatment: How far away from your imagined trend line are most of the data points?
- Very near - 1 2 3 - Very far

- Is there a converging difference in variability between Treatment and Baseline? Yes (No)
  - Baseline: variability rating = 1
  - Treatment: variability rating = 1
  - No, they are the same
Step 4: Make a Decision

Summarize
Was there a convincing change in level? Yes / No

- Baseline y-axis value = 7
- Treatment y-axis value = 3

Step 4: Make a Decision

Summarize
Was there a convincing change in trend? Yes / No

- Baseline descending (down)
- Treatment descending (down)
- Same steepness

Step 4: Make a Decision

Summarize
Was there a convincing change in variability? Yes / No

- Baseline very near — 2 3 — Very far
- Treatment very near — 2 3 — Very far

If you answered "No" for ALL of these, skip the Level Exception section and answer NO in (c) below

We did not answer "No" for all of these, so we will complete the Level Exception section

Step 4: Make a Decision

LEVEL EXCEPTION

If level was the only dimension that changed, AND the trend was the same in Baseline and Treatment (ascending or descending, NOT flat), it is a Level Exception graph.

The questions on the next slides and on your job aid will help you determine if the level exception applies.

These are level exception graphs:

If you remove the phase line, a level exception graph will look like the data paths could be part of the same line, like the graph on the right
Step 4: Make a Decision

1. Was the only other axis that changed?  
   - No, it was the same.  
   - Yes, it changed.  

2. Level measure for the level in baseline and treatment?  
   - Yes, it was the same.  
   - No, it was different.  

3. Did trend measure for the level in baseline and treatment?  
   - Yes, it was the same.  
   - No, it was different.  

4. Did behavior change from Baseline to Treatment?  
   - Yes, level changed.  
   - No, level did not change.

If YES in 1, 2, 3 and 4 answer NO in (b) below:

(c) Did behavior change from Baseline to Treatment?  
   - Yes  
   - No  

Our Example

- Did behavior change from Baseline to Treatment?  
  - Yes  

Because level was the only dimension that changed, and the trend was the same, it is a Level Exception graph.

Let's do one more...

Here is our graph

Step 1: Level

Baseline, What is the mean level?  
5.5

Step 1: Level

Treatment, What is the mean level?  
5.5

Step 1: Level

Compare the last few points of baseline to the first few points of treatment. Was there an immediate change in level between phases?

No, the level is approximately the same.
Step 1: Level

Is there a difference between the levels (y-axis values) of the data points in Treatment and Baseline? Yes (No)
- Baseline: y-axis value = 5.5
- Treatment: y-axis value = 5.5

No, they are the same.

Step 2: Trend

Is the trend steeper, flatter, or neither steeper nor flatter in Treatment than in Baseline?
- steeper
- flatter
- neither

The trend is flat in both treatment and baseline.

Step 2: Trend

Is there a change in direction of trend from Baseline to Treatment? Yes (No)
- Baseline: No trend (flat)
- Treatment: No trend (flat)

No, they are the same.

Is there a change in slope (steepness) of the trend in Treatment compared to Baseline? Yes (No)
- Treatment: Neither steeper nor flatter

No, there is no change in slope.

Step 3: Variability

How far away from your imagined trend line are most of the data points? Very near - 1 2 3 - Very far
Step 3: Variability

Is there a convincing difference in variability between Treatment and Baseline? Yes (No)
- Baseline: variability rating = 3
- Treatment: variability rating = 3

No, they are the same.

Step 4: Make a Decision

Summarize
Was there a convincing change in level? Yes (No)
Baseline: y-axis value = 5.5
Treatment: y-axis value = 5.5

Was there a convincing change in trend? Yes (No)
Baseline: No trend (flat)
Treatment: No trend (flat)

Was there a convincing change in variability? Yes (No)
Baseline: Very Near = 1 2 3 - Very Far
Treatment: Very Near = 1 2 3 - Very Far

If you answered “No” for ALL of these, skip the Level Exception section and answer NO in (c) below
Step 4: Make a Decision

LEVEL EXCEPTION

1) Was level the only dimension that changed?  
   Yes (a), No, answer YES in (b) below.
2) Was trend the only dimension that changed?  
   Yes (a), No, answer YES in (b) below.
3) Was variability the only dimension that changed?  
   Yes (a), No, answer YES in (b) below.
4) Did behavior change from Baseline to Treatment?  
   Yes (a), No, answer YES in (b) below.

If YES in 1, 2, 3, and 4, answer NO in (c) below.

Our Example

- Was there a convincing change in level? Yes (a), No
- Was there a convincing change in trend? Yes (a), No
- Was there a convincing change in variability? Yes (a), No

Did behavior change from Baseline to Treatment?  Yes (a), No

Because there were no changes in level, trend, or variability.

Important Points

- Remember to fill out all of your job-aids completely as you analyze your graphs.
  - You will use one job-aid per graph (10 job-aids for each packet of graphs).
  - Tell the researcher if you need more job-aids.
- If you disagree with the decision reached by using the job-aid, circle the decision obtained with the job-aid anyway (not your opinion).
  - We want to know how well the job-aid works, not how well you analyze graphs without it.

Questions

- If you have questions about how to use the job-aid, please ask them now.
- You may also ask questions as you fill out your graph packets.
- Please keep in mind that the researcher may not be able to answer your questions.
  - Specifically, the researcher will not answer questions about whether your decisions are right or wrong.
  - If the researcher cannot answer your question, he or she will tell you to just do the best you can. (We know this is frustrating - sorry!)
Appendix C

Faded Job-Aids

Step 1: Level

Baseline:
• What y-axis number does the level line cross? 

Treatment:
• What y-axis number does the level line cross? 

☑ Is the level (y-axis value) of the data path in Treatment different than in Baseline? Yes / No

Step 2: Trend

Baseline:
• What is the trend? ascending (up)  descending (down)  no trend (flat)

Treatment:
• What is the trend? ascending (up)  descending (down)  no trend (flat)

☑ Is the direction of the trend in Treatment different than in Baseline? Yes / No

Step 3: Variability

Baseline: How far away from your imagined trend line are most of the data points?
Very near -- 1 2 3 4 5 -- Very far

Treatment: How far away from your imagined trend line are most of the data points?
Very near -- 1 2 3 4 5 -- Very far

☑ Is the variability in Treatment different than in Baseline? Yes / No

Step 4: Make a Decision

a. Summarize
  How much did level change? no change a little a lot
  How much did trend change? no change a little a lot
  How much did variability change? no change a little a lot

  • If you answered “no change” for ALL of these, answer NO in (c) below and skip the Level Exception section.

b. Level Exception
  1. Was level the only dimension that changed? Yes / No
     If NO, answer YES to (c) below
  2. Did the trend stay the same in Baseline and Treatment? Yes / No
     If NO, answer YES to (c) below
  3. Was trend ascending or descending in both Baseline and Treatment? Yes / No
     If NO, answer YES to (c) below

  • If you answered Yes to questions 1, 2 AND 3 above, answer No in (c) below.

c. ☑ Did behavior change from Baseline to Treatment? Yes / No
Visual Inspection Job-Aid

☑ Is the level (y-axis value) of the data path in Treatment different than in Baseline? Yes / No

☑ Is the trend in Treatment different than in Baseline? Yes / No

☑ Is the variability in Treatment different than in Baseline? Yes / No

☑ Is this a Level Exception graph? Yes / No

☑ Did behavior change from Baseline to Treatment? Yes / No
Appendix D

Traditional Lecture Outline

From Cooper, Heron, & Heward (2007)

Purpose of Visual Analysis

*Systematic means of evaluating behavior-analytic data*

Seeks to determine:

- Was there a meaningful change in behavior?
- To what extent can the change in behavior be attributed to the independent variable?

Process of Visual Inspection

Determine whether the graph is fit to be analyzed

Read axis labels, legend, phase labels

Examine scaling of axes, including scale breaks

Identify what each data point represents

Raw scores from single observations vs. averages/some type of summary from multiple observations

Performance of one subject vs. group of subjects

If averages/summaries, are ranges/variation depicted?

Are the data accurately represented by the display?

Inspect the data within conditions

Number of data points

The more observations over the longer period of time, the more confident one can be that the sample represents the true course of behavior change

Fewer points are needed in subsequent replications if the level, trend, variability is similar to that in previous phases of that condition
More data points are needed to demonstrate new findings

Exceptions:

It is unethical to perform multiple observations of dangerous behavior such as SIB under conditions in which there is no reasonable expectation for improvement (e.g., no-treatment baseline)

It is not helpful to conduct repeated measurements when the subject cannot logically perform the response or when there is no opportunity for the behavior to occur

Variability

Variability is the frequency and extent to which repeated measures of behavior yield different outcomes

A lot of variability or a high degree of variability suggests poor control of the factors influencing the behavior

More data points are needed to establish a predictable pattern of performance when variability is high; fewer are needed when data show little variability

Level

Level is the vertical-axis value around which repeated measures of behavior converge

Level must be considered with regard to the level of variability in the data (e.g., a mean level line may not be representative of any of the behavioral measurements in the phase, such as when performance is initially high and stable but then low and variable within a phase)

Trend

Trend is the overall direction taken by a data path

Trends are described in terms of their:

Direction (increasing, decreasing, zero)

Degree/magnitude
Variability of data points around the trend

The trend of a series of data points can be represented by a straight line drawn through the data (i.e., trend line or line of progress)

- **Freehand** (visual estimate; ignore 1-2 extreme outliers)
- **Least-squares trend line** (computed using the ordinary least-squares linear regression equation)
- **Split-middle line of progress**
  
  **Step 1:** Divide the data into two equal parts
  
  **Step 2:** Find the intersections of the mid-rate and mid-date for each half
  
  **Step 3:** Draw a line that passes through both of the intersections
  
  **Step 4:** Move the line up or down (keeping it parallel to the original line) such that the same number of data points fall on and above the line as fall on and below the line
Inspect the data between conditions

Level

Look at the last data point before the condition line and the first data point after the condition line to determine whether there was an immediate change in behavior with the manipulation of the independent variable

Compare the overall level of performance between conditions (consider overlap and delayed or temporary effects)

Trend

Compare changes in direction or slope of trend between phases

Examine performance across not only adjacent phases, but also across similar conditions

Evaluate the experimental design to determine whether the change can be attributed to the independent variable
Appendix E

Traditional Lecture PowerPoint Presentation

Visual Inspection

Analyzing Single-Case Design Data

Purpose of Visual Analysis

- Systematic means of evaluating behavior-analytic data
- Seeks to determine:
  - Was there a meaningful change in behavior?
  - To what extent can the change in behavior be attributed to the independent variable?

Process of Visual Inspection

- Determine whether the graph is fit to be analyzed
- Inspect the data within conditions
- Inspect the data between conditions
- Evaluate the experimental design to determine whether the change can be attributed to the independent variable

Determine Whether the Graph is Fit to Be Analyzed

- Read all labels
- Examine scaling of axes
- Identify what each data point represents
- Decide whether the data are accurately represented by the display

Inspect the Data Within Conditions

- Number of data points
  - More observations lead to more confidence that the sample is representative of the behavior
  - Fewer data points are needed in subsequent replications
  - More data points are needed to demonstrate new findings
  - Exceptions

Variability
- The frequency and extent to which repeated measures of behavior yield different outcomes
- A lot of variability or a high degree of variability suggests poor control of the factors influencing the behavior
- More data points are needed to establish a predictable pattern of performance when variability is high
Variability Examples

Level

- The vertical-axis value around which repeated measures of behavior converge.
- Must be considered with regard to the level of variability in the data.
  - A mean level line may not be representative of any of the behavioral measurements in the phase, such as when performance is initially high and stable but then low and variable.

Level Examples

Mean Level Line is Not Representative

Inspect the Data Within Conditions

- Trend
  - The overall direction taken by a data path.
  - Described in terms of:
    - Direction (increasing, decreasing, zero)
    - Degree/magnitude (slope)
  - Variability of data points around the trend.
  - Can be represented by a straight line drawn through the data (trend line or line of progress).

Trend Examples
Drawing Trend Lines

- Freehand
  - Visual estimate
  - Ignore 1-2 extreme outliers
- Least-squares trend line
  - Uses mathematical equation: ordinary least-squares linear regression equation
  - Hard to do without a computer

Split-Middle Line of Progress

- Condition/phase line indicates independent variable manipulation
- Level
  - Look for an immediate change in level
  - Compare overall level of performance between conditions
  - Consider overlap
  - Look for delayed or temporary effects

Level Change Examples

Inspect the Data Between Conditions

- Trend
  - Changes in direction
  - Changes in slope
- Examine performance across adjacent phases and across similar conditions
### Trend Change Examples

![Graph showing trend change examples]

### Evaluate the Experimental Design

- Determine whether the change in behavior can be attributed to the independent variable.
Appendix F

Fidelity Checklist for Traditional Lecture Video

Rater: ____________________

*Please circle Yes or No to indicate whether the following components were included in the video lecture.*

1. A rationale for the visual analysis of graphs? Yes / No
2. A description of the following concepts?
   a. Level Yes / No
   b. Trend Yes / No
   c. Variability Yes / No
3. Information on determining whether a graph is fit to be analyzed? Yes / No
4. Information on inspecting data within conditions? Yes / No
5. Instruction on drawing trend lines? Yes / No
6. Information on inspecting data across conditions? Yes / No
7. A statement about evaluating experimental design? Yes / No
8. Example graphs of the following phenomena?
   a. Little variability Yes / No
   b. High variability Yes / No
   c. Estimation of mean level Yes / No
   d. Mean level that is unrepresentative of performance during the behavioral measurements Yes / No
   e. Increasing trend Yes / No
   f. Decreasing trend Yes / No
g. Zero trend  
  Yes / No

h. Split-middle line of progress  
  Yes / No

i. Change in level between phases  
  Yes / No

j. Change in trend between phases  
  Yes / No

9. Overall, does the video provide a description of visual analysis consistent with common text recommendations?  
   Yes / No
Appendix G

Graph Analysis Concept Quiz

Participant: __________

1. What is the primary method of data analysis in single-case experimental designs?
   a. Direct observation of the client to see if behavior has changed
   b. Visual inspection of graphed data
   c. Statistical analysis of raw data
   d. Visual inspection of graphed data with confirmation by statistical tests

2. What does a phase line on a graph represent?
   a. The passage of time
   b. The number of sessions
   c. A manipulation of the independent variable
   d. A different group of participants

3. Trend can change in __________ and/or __________.
   a. Level; variability
   b. Number; extent
   c. Phase; data
   d. Direction; slope

4. The process of visual inspection includes all of the following except:
   a. Determining who conducted the study and if they are good researchers
   b. Determining whether the graph is fit to be analyzed
   c. Inspecting the data within conditions
   d. Inspecting the data across conditions
   e. Evaluating the experimental design to determine whether the change can be attributed to the independent variable

5. The frequency and extent to which repeated measures of behavior yield different outcomes is referred to as:
   a. Trend
   b. Level
   c. Variability
   d. Spread

6. Which of the following factors is most important in determining whether a graph demonstrates an effect?
   a. Level
   b. Trend
   c. Variability
   d. All of these are important and must be considered in combination

7. The average amount of behavior in a phase is referred to as:
   a. Variability
   b. Level
   c. Trend
   d. Spread
Appendix H

HSIRB Approval Letter

Date: May 11, 2009

To: Wayne Fuqua, Principal Investigator
    Candice Jostad, Student Investigator for dissertation

From: Amy Maugle, Ph.D., Chair

Re: Extension and Changes to HSIRB Project Number 08-05-08

This letter will serve as confirmation that the extension and changes to (remove traditional-tenure-only control group) your research project "Teaching Visual Inspection Skills to College Students" requested in your memo received May 8, 2009 have been approved by the Human Subjects Institutional Review Board. The conditions and the duration of this approval are specified in the Policies of Western Michigan University.

Please note that you may only conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition, if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: May 12, 2010