FURTHER EVALUATION OF A DECISION-MAKING ALGORITHM TO AID IN TRAINING VISUAL ANALYSIS

By

Darla Kril

A THESIS

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Applied Behavior Analysis – Master of Arts

2019
ABSTRACT

FURTHER EVALUATION OF A DECISION-MAKING ALGORITHM TO AID IN TRAINING VISUAL ANALYSIS

By

Darla Kril

Visual analysis is an important component of evaluating graphically displayed data and making clinical decisions in Applied Behavior Analysis (ABA). The individuals responsible for implementing behavioral interventions and analyzing data are often behavior technicians who may not be provided with effective training to be proficient in visually analyzing data. Therefore, there is a need for an effective and streamlined method to train visual analysis. Previous research has suggested the support of a clinical decision-making algorithm (DMA) to aid in making correct clinical judgments. The current study trained six graduate students to make clinical decisions using a computer-based training and the aid of the DMA. Correct responding was evaluated upon removal of the DMA. Five of the six participants increased their percentage of accurate responses with the DMA alone. High levels of correct responding maintained for four of the six participants.

Keywords: decision-making algorithm, feedback, replication, visual analysis
TABLE OF CONTENTS

LIST OF FIGURES .............................................................................................................ii

INTRODUCTION ................................................................................................................1

METHOD ...............................................................................................................................4
  Participants and Setting .................................................................................................4
  Materials .........................................................................................................................4
    Decision-making algorithm .........................................................................................4
    Graphs and module development ..............................................................................4
  Dependent Measure .......................................................................................................6
    Percentage of correct responses ................................................................................6
  Experimental Design and Procedure ..........................................................................6
  Procedure ......................................................................................................................6
    Baseline ......................................................................................................................7
      Decision-making algorithm ......................................................................................8
    Feedback session (Participant 4 only) ......................................................................8
    Return to baseline .....................................................................................................8
    Two-week follow-up probe .......................................................................................9

RESULTS ...........................................................................................................................10
  Participant 1 ..................................................................................................................10
  Participant 2 ..................................................................................................................10
  Participant 3 ..................................................................................................................10
  Participant 4 ..................................................................................................................11
  Participant 5 ..................................................................................................................11
  Participant 6 ..................................................................................................................12

DISCUSSION ....................................................................................................................13

APPENDIX .......................................................................................................................18

REFERENCES ..................................................................................................................19
LIST OF FIGURES

Figure 1. Flow chart: The decision-making algorithm is used to determine which clinical decision to make.................................................................19

Figure 2. Graph: Results for Participants 1, 2, and 3. The * denotes the two-week follow-up probe.................................................................20

Figure 3. Graph: Results for Participants 6, 4, and 5. The * denotes the two-week follow-up probe. The # denotes the session completed immediately following the feedback session. The results of the feedback session are not displayed above.................................21
INTRODUCTION

Applied behavior analysis (ABA) uses behavioral principles to generate socially significant behavior change (Baer, Wolfe, & Risley, 1968). The variables responsible for behavior changes are strategically isolated and manipulated to create and evaluate behavioral treatments (Cooper, Heron, & Heward, 2007). To determine if a treatment is effective, data are usually recorded, then presented and analyzed in a graphic format. This method of treatment evaluation is known as visual analysis (Kazdin, 2011).

There are several ways to visually analyze data. Common methods include analyzing trend (i.e., the direction of a data path), level (i.e., the absolute value on the vertical axis, or stability), and variability (i.e., the frequency and extent to which the same measure yields different results; Cooper et al., 2007). If comparing two or more experimental conditions, visual analysis methods may also include a determination of the immediacy of effect (i.e., the magnitude of effect between conditions; Kratochwill & Levin, 1992). Aside from visual analysis, statistical evaluations, such as percentage of non-overlapping data (Scruggs & Mastropieri, 2013), and Tau-U (Kratochwill & Levin, 1992) may also aid in the identification of a treatment effect.

Because visual analysis is a critical component in ABA interventions, it is necessary to empirically test methods to train front-line employees to demonstrate this skill (Kratochwill et al., 2014). Previous studies have examined treatment packages to teach visual analysis, using visual aids (Fisher, Kelley, & Lomas, 2003), video instruction (Stewart, Carr, Brandt, & McHenry, 2007), within stimulus prompting (Young & Daly, 2015), and written rules (Maffei-Almodovar, Feliciano, Fienup, & Sturmey, 2017).
There are several limitations mentioned to the treatment packages described. The first limitation is that the studies evaluated methodology to identify the degree of treatment effect by comparing two conditions. Though this analysis is important, practitioners of ABA interventions may not always compare two conditions, and instead evaluate data in one continuous line-series graph. A second limitation is that the previously mentioned studies either consisted of trainings that were long in duration (e.g. Eight and twelve minute training videos followed by analyzing a variety of graphs; Stewart et al., 2007) or required the physical presence of a trainer or interventionist, which can be costly (Wolfe & Slocum, 2015). Therefore, it is necessary to develop cost effective and efficient methods to train visual analysis to employees who directly implement ABA treatments.

Kipfmiller et al. (2019) addressed the above limitations of previous visual analysis research by evaluating the effects of training visual analysis skills to newly-hired behavior technicians at a community-based autism center using a clinical decision-making model (CDMM). The CDMM consisted of four clinical decisions: (a) continue intervention, (b) discontinue intervention, (c) modify intervention, and (d) intervention is complete. The CDMM aided participants in answering questions regarding a set of hypothetical data in the form of a line graph and answered a series of predetermined questions in sequential order to analyze data. When the CDMM was present, accuracy of visual analysis of graphs increased. In-person feedback sessions were incorporated for two participants when correct responding did not increase.

The increase in performance suggested the CDMM was effective in training visual analysis for six of the eight participants. This provided a potentially efficient and cost-effective method to training front line employees in visual analysis. One limitation is that the CDMM was
not removed during the study. Therefore, it is unclear if exposure to the CDMM may serve as an effective instructional tool for training visual analysis of time-series graphs. Furthermore, maintenance of trained visual analysis skills was not evaluated, therefore the long-term effects of exposure to the CDMM in interpreting time-series graphs are unknown. For the purpose of this study, the CDMM was refined and is hereafter referred to as the decision-making algorithm (DMA; Karston, et al., 2011)

The purpose of the current study is to extend the findings in Kipfmiller et al. (2019) by evaluating if correct responding continues to occur upon the removal of the DMA. A second purpose of this study was to evaluate accuracy of visual analysis during a two-week follow up period to evaluate short-term maintenance of visual analysis skills. Finally, a third purpose of this study was to provide electronic feedback, instead of in-person feedback (as reported in Kipfmiller et al.), to further reduce cost and effort associated with teaching participants to use the DMA.
METHOD

Participants and Setting

Six female participants (age range: 20 to 26) were recruited to participate in this study. Participants were students in a graduate course in behavior analysis at a large public university in the Midwest. All participants had prior experience as a behavior technician. All but one participant (Participant 5) had previous experience in behavior analysis and visual analysis. Upon providing consent to participate in the study, participants were separated into two groups to create two multiple baseline designs. These groups were created based on the order of participants’ completion of Module 1 (described below).

Materials

Decision-making algorithm. The DMA was a modified version of that reported in Kipfmiller et al. (2019). Modifications included the rearrangement of the placement of arrows and boxes in the DMA graphic. This was altered to increase the clarity of pathways in the DMA for participants. The DMA was depicted as a flow chart that asks a series of yes or no questions. The user of the DMA is then guided to one of four specific clinical recommendations (i.e., continue intervention, discontinue intervention, modify intervention, or intervention is complete) based on the answers they provided (see Figure 1).

Graphs and module development. Graphs were constructed and sorted into four categories (i.e., continue intervention, discontinue intervention, modify intervention, or intervention is complete) using the procedures described by Kipfmiller et al. (2019). Graphs consisted of 10 data points. Data points represented hypothetical 10-trial blocks of data. On each graph, the x-axis was labeled “sessions” and the y-axis was labeled “percentage of correct responses”. Gridlines at 20% intervals were present to allow for ease of interpretation. No phase
change lines or additional information were present. Sixty graphs were created for each category, for a total of 240 unique graphs.

Prior to beginning the experiment, two modules (quizzes) were created on Desire2Learn (D2L) with a third module created during the experiment. D2L served as the interface where participants’ visual analysis skills were evaluated. Module 1 consisted of 12 questions. Each question contained one randomly selected graph, from the pool of 60 available graphs, from one of the four categories above. Each question also included the statement “Please select the correct answer”, and the options (a) Continue Intervention, (b) Discontinue Intervention, (c) Modify Intervention, and (d) Intervention is Complete were presented in a random order. Only one question was available at a time, and participants could not return to previous questions once an answer was selected. Participants were able to select only one answer per question. A total of 12 graphs were randomly selected (three graphs from each of the four categories). Module 1 was programmed to randomly generate graphs from each category each time it was accessed to ensure each module was unique every time it was accessed. There was no time limit to complete each question.

Module 2 was identical to Module 1, with the exception that participants were presented with the DMA alongside each question within the module (described in more detail below). An additional passage was included instructing participants to use the support of the DMA to respond to questions in the module.

A third module was created that was identical to Module 1, with one modification. Specifically, this module provided computer-based feedback to participants regarding the accuracy of their responses for each question. Participants were able to view their response, along with the correct response upon completion of the Module. Participants were only assigned
this module if responding upon introduction of the DMA did not yield an increase in percentage of accurate responses.

A list of recommended web-browsers was provided in the instructions for participants to ensure optimal graph and question display. These browsers were tested prior to beginning the study to confirm instructions, responses, the graph, and the DMA appeared on each screen.

**Dependent Measure**

**Percentage of correct responses.** The percentage of participants’ correct responses was the dependent measure. A correct response was defined as any instance when a participant’s answer to a quiz question matched the correct clinical decision that corresponded to the graph depicted in that question. Percentage of correct responses was calculated by dividing the number of correct responses for each session, by the total number of questions (i.e. twelve), and then multiplying it by 100 to yield a percentage (Cooper et al., 2007).

**Experimental Design and Procedure**

To evaluate the effectiveness of training visual analysis using the DMA, a withdrawal embedded in a non-concurrent multiple baseline across participants design was used. This design allows for the demonstration of experimental control by systematic replication of effects across all participants. Staggering the introduction of the intervention across the participants controls for history throughout the study (Ledford & Gast, 2018). A withdrawal is embedded within this design to evaluate if participants require the DMA to be present to accurately analyze graphic data or if visual analysis skills are acquired after brief exposure to the DMA.

**Procedure**

All experimental sessions and conditions were completed on each participant’s personal computer. An experimenter was not present during sessions. Prior to start of the study,
participants were asked to download a recommended web-browser: Google Chrome, Microsoft Edge, Safari, or Firefox. Participants were then provided with instructions to log into D2L. Participants received a document with detailed instructions on how to create a username and password on D2L. Once participants created a username and password, a message was sent by the experimenter through D2L with instructions to complete a pre-determined number of sessions of baseline Participants completed sessions

**Baseline.** The purpose of baseline was to establish a steady-state of participant responding prior to the introduction of the DMA. Participants were asked to complete Module 1 five times to ensure level, trend, and variability could be determined (Kratochwill et al., 2010)

Before each baseline session, each participant was provided with the following instructions:

> In this module, you will be shown a series of graphs and asked to select between 4 possible clinical decisions, based on information in that graph. Please answer each question to the best of your ability. You may not use outside materials, and you may not consult with your peers. Please complete the quiz on your own.

Participants then selected “Start quiz” to begin. The first question included the instruction “Please select the correct answer”, four clinical decision-making outcomes, and a graph. The DMA was not present. Participants then selected an answer and clicked “Next” to move to the next question. There was no opportunity to return to previous questions and no feedback was provided at any point during baseline.

To minimize testing as a threat to internal validity, Module 1 was programed to randomly select 12 graphs from each of the four categories, for a total of three randomly selected graphs from each category. Each category contained 60 unique graphs as to minimize the probability of the same graph presented multiple times during the study.
**Decision-making algorithm.** After completing Module 1, Participants were asked to complete Module 2 within one week. Sessions could be completed immediately following one another or spaced across days. The purpose of this condition was to evaluate each participant’s response accuracy in the presence of the DMA. Module 2 was identical to Module 1, with the following modifications. First, The DMA was provided directly above each graph for each question. Second, in addition to the instructions provided in Module 1, the following passage was added to the instructions the participants viewed each time they accessed Module 2: “With each graph, we have provided you with a visual aid. Please use the visual aid to answer each question to the best of your ability.”

There was no time limit for these sessions and participants did not receive feedback on any answer during this condition. Participants were required to complete at least five sessions and reach stable responding at 92-100% to continue to the return to baseline condition. This provides a high level of responding prior to removal of the model.

**Feedback session (Participant 4 only).** A feedback session was administered if the percentage of correct responses did not increase above baseline levels during the decision-making algorithm condition. Module 3 was identical to Module 2 with the exception that the participant was able to view their answer, along with the correct answer, following the completion of the module.

**Return to baseline.** The purpose of this condition was to evaluate the percentage of correct responses for each participant following the withdrawal of the DMA. This condition was identical to that of baseline. The DMA was not present. No feedback was provided to participants at any time.
Two-week follow-up probe. The purpose of this condition was to evaluate participant responding after the passage of two weeks since their last exposure to baseline conditions. This condition was identical to that of baseline.
RESULTS

The results of participants’ percentage of correct responses are depicted in Figures 2 and 3. Overall, the results indicated five of the six participants demonstrated an increase in correct responding when the DMA was introduced. One participant required a feedback session during intervention. Correct responding maintained for five of the six participants when the DMA was removed. A two-week follow-up probe was conducted and performance maintained for four of the six participants.

Participant 1

During baseline, Participant 1 averaged 48% (range: 33-75%) correct responding with a downward trend across five sessions (Figure 2, top panel). Upon introduction of the DMA, an immediate treatment effect was observed (increasing from 33% to 100%). A steady trend of responding occurred across five sessions with an average of 98% (range: 92-100%) correct responding. After the DMA was removed, responding was at 93% (range: 83-100%) with minimal variability. At a two-week follow-up, responding was 17%.

Participant 2

During baseline, Participant 2 averaged 74% (range: 58-92%) correct responding and responding was highly variable across seven sessions (Figure 2, middle panel). Upon introduction of the DMA, an immediate treatment effect was observed (increasing from 67% to 100%). A steady trend of responding occurred across seven sessions with an average of 98% (range: 92-100%) correct responding with little variability. After the DMA was removed, correct responding occurred at an average of 98% (range: 92-100%) with minimal variability. At a two-week follow-up, responding was at 100%.

Participant 3
During baseline, Participant 3 averaged 57% (range: 25-83%) correct responding with high variability across nine sessions (Figure 2, bottom panel). Upon introduction of the DMA, an immediate increase in correct responding occurred. A steady trend of responding occurred across six sessions with an average of 93% (range: 92-100%) correct responding. After the DMA was removed, correct responding occurred at an average of 89% (range: 75-100%) with moderate variability across seven sessions. At a two-week follow-up, responding was at 92%.

**Participant 4**

During baseline, Participant 4 averaged 42% (range: 33-50%) correct responding across five sessions (Figure 3, top panel). Upon introduction of the DMA, an initial increase in responding occurred. A downward trend then occurred averaging 47% (range: 25-67%) across five sessions. One computer-based feedback session was provided to the participant. Correct responding occurred at 50% during the feedback session (i.e. Six correct responses out of twelve questions). Correct responding immediately increased to an average of 98% (range: 92-100%) across five sessions, following the feedback session. After the DMA was removed, correct responding occurred at 100% across five sessions. At a two-week follow-up, responding was at 100%.

**Participant 5**

During baseline, Participant 5 averaged 34% (range: 25-50%) correct responding with a variable trend across seven sessions (Figure 3, middle panel). Upon introduction of the DMA, an immediate treatment effect was observed (increasing from 33% to 100%). A steady trend of responding occurred across five sessions with an average of 98% (range: 92-100%) correct responding. After the DMA was removed, correct responding immediately decreased to 50% across five sessions. At a two-week follow-up, responding was at 33%.
Participant 6

During baseline, Participant 6 averaged 59% (range: 33-83%) correct responding with high variability across nine sessions (Figure 3, bottom panel). Upon introduction of the DMA, an immediate treatment effect was observed (increasing from 58% to 100%). A steady trend of responding occurred across five sessions with an average of 98% (range: 92-100%) correct responding. After the DMA was removed, responding was at an average of 95% (range: 83-100%) with minimal variability. At a two-week follow-up, responding was at 100%.
DISCUSSION

Following baseline, percentage of correct responses increased for five of the six participants when the DMA was present; and one participant (Participant 4) required a feedback session to increase percentage of correct responses. Five of the six participants maintained high levels of percentage of correct responding that was above baseline levels once the DMA was removed. During 2-week follow-up probes, high levels of correct responding maintained for four of the six participants. Further discussion of the implications of these results are discussed below.

Participants were asked to complete five baseline sessions. While participants were given a time frame in which these sessions must be completed (less than one week), it was not specified if sessions were to be completed in rapid succession or spaced across days. All participants, apart from Participant 5 and 6, completed the five assigned baseline sessions in rapid succession. Participant five completed two sessions subsequently and the following day completed three sessions subsequently. Participant 6 completed one session and completed four of five sessions in rapid succession two days later. Participants 1 and 6 were the only two participants assigned Module 2 following only five baseline sessions. Participants 1-6 took an average of 3.54, 2.55, 3.42, 3.9, 3.04, and 2.08 minutes, respectively, to complete five baseline sessions. The majority of participants continued to complete sessions quickly and in rapid succession. While modules could be completed quickly, future research could evaluate whether completing tasks in rapid succession or spacing tasks out lead to higher accuracy of responding during tasks (Donovan & Radosevich, 1999).

Upon introduction of the DMA, accuracy immediately increased for five of the six participants. However, participant 4 demonstrated a downward trend in the data path across five sessions. The reason for this decrease in correct responding (range: 25-67%) is unclear.
Participant 4 was exposed to a single session (i.e., 12 questions) of Module 3 where she was provided with computer-generated feedback for all 12 questions and answers at the end of that quiz. This feedback session served as a methodological refinement of Kipfmiller et al. (2019), who used in-person feedback when participant responding did not increase in the presence of the DMA. Following exposure to one feedback session, Participant 4’s percentage of correct responding immediately increased. Participant 4 continued to engage in high levels of responding (range: 92-100%) across the remainder of the experiment. This finding suggests administering computer-generated feedback may be an efficient way to increase accurate responding and promote maintenance. Another implication of these findings includes the possible benefits of training professionals on using a decision-making algorithm prior to its implementation. These present important areas for future research.

Upon removal of the DMA, percentage of correct responses maintained above baseline levels for five of the six participants. However, responding was more variable than during the decision-making algorithm condition for three participants (Participants 1, 3, and 6). Visual analysis is an important skill to demonstrate with high levels of accuracy as it drives treatment decisions. An implication of this is incorrect clinical decisions may significantly affect treatment outcomes similar to the effect of treatment efficacy of interventions on treatment outcomes (Carroll, Kodak, & Fisher, 2013). It is possible this could be a predictor, among others, of accuracy of responding to visual analysis tasks and should be considered a possible direction for future research.

The latency between completion of Module 2 and returning to Module 1 varied among participants. Participants 1 and 2 completed Module 1 (return to baseline) one day after completing Module 2. Participant 4 completed Module 1 (return to baseline) two days after, and
two participants (Participant 3 and 6) completed Module 1 (return to baseline) three days following Module 2. Participant 5 did not complete Module 1 (return to baseline) until six days after the completion of Module 2. Upon removal of the DMA in return to baseline, Participant 4 engaged in low levels of accurate responding (50% across five sessions). It is possible this is due to the latency between being exposed to the DMA and its removal, making this an area for potential future research.

In addition, at a two-week follow-up probe, four of the six participants maintained high percentages of correct responding (Participants 2, 3, 4, and 6). Two participants engaged in lower percentages of correct responding than in the previous condition (17% and 33% for Participants 1 and 5, respectively). These findings suggest that although the DMA may increase participant accuracy when present, exposure to the DMA may not serve as an adequate tool to teach decision-making, at least in the context of our study. The differences in outcomes across participants is unclear and presents an obvious avenue for future research.

There are at least a few limitations of this study that are worth noting. One limitation is the limited number of participants ($n = 6$), as well as the participant pool itself. The participants were comprised of six graduate students with minimal experience in visual analysis and ABA and experience as behavior technicians. This group may not be representative of all front-line employees who may visually analyze data in treatment settings. Another limitation is that we used hypothetical graphs, each with 10 data points, that were developed using an auto-regressive formula. These graphs were used in order to control for task difficulty and practice effects within and across sessions, and do not resemble what practitioners will likely encounter in the real world. Future research should evaluate the effects of the DMA in analyzing graphs on an ongoing basis in order to further evaluate its applied utility.
Previous scholarship that has described a decision-making algorithm has also provided a description of how to use that algorithm (e.g., Karsten et al., 2011). However, no specific instructions were provided to our participants. This omission is a limitation and presents an important area for future research. Future research could evaluate the effects of a brief instructional session when the DMA is introduced to increase the probability participants use it as intended. It is possible that Participant 4 may have engaged in higher percentage of correct responses, upon introduction of the DMA, if she had been taught how to use it.

In addition to an instructional session prior to the introduction of the DMA, future research should also continue to refine feedback procedures if the DMA does not improve participant responding. Data from Participant 4 presents preliminary results of a more efficient and effective modality of feedback than what was reported in Kipfmiller et al. (2019). Though our experimental tasks were identical to those described by Kipfmiller et al., only one participant was exposed to our feedback session. Therefore, this finding should be taken with caution.

Participants were exposed to the DMA for a varying number of sessions (range: 5-10 sessions). One additional area for future research may be to increase length of exposure to the DMA, by increasing the number of sessions each participant is exposed to the DMA, to evaluate if length of exposure may increase the probability of accurate responding once the DMA is removed. A multi-element design may present an ideal arrangement to evaluate this particular question. With the exception of the feedback session provided to one participant during the intervention phase, five of the six participants achieved accurate and stable responding with the presence of the model. Five of the six participants maintained accurate responding once the DMA was removed. At a two-week follow-up probe correct responding maintained for four of
the six participants. While the DMA may serve as a supplementary tool, further research is needed to determine an efficient and low-cost primary method to train visual analysis.
APPENDIX
Figure 1. Flow chart: The decision-making algorithm is used to determine which clinical decision to make.
Figure 2. Graph: Results for Participants 1, 2, and 3. The * denotes the two-week follow-up probe.
Figure 3. Graph: Results for Participants 6, 4, and 5. The * denotes the two-week follow-up probe. The # denotes the session completed immediately following the feedback session. The results of the feedback session are not displayed above.
REFERENCES
REFERENCES


