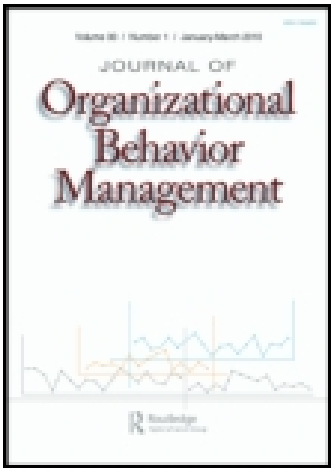

Journal of Organizational Behavior Management

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/worg20

The Effects of Evaluating Video Examples of Staffs' Own Versus Others' Performance on Discrete-Trial Training Skills in a Human Service Setting

W. Larry Williams [a] & Julianne Gallinat [a]

[a] University of Nevada, Reno , Reno, Nevada, USA

Published online: 20 May 2011.

To cite this article: W. Larry Williams & Julianne Gallinat (2011) The Effects of Evaluating Video Examples of Staffs' Own Versus Others' Performance on Discrete-Trial Training Skills in a Human Service Setting, Journal of Organizational Behavior Management, 31:2, 97-116, DOI: 10.1080/01608061.2011.570099

To link to this article: http://dx.doi.org/10.1080/01608061.2011.570099

Journal of Organizational Behavior Management, 31:97–116, 2011

ISSN: 0160-8061 print/1540-8604 online
DOI: 10.1080/01608061.2011.570099

RESEARCH ARTICLE

The Effects of Evaluating Video Examples of Staffs' Own Versus Others' Performance on Discrete-Trial Training Skills in a Human Service Setting

W. LARRY WILLIAMS and JULIANNE GALLINAT
University of Nevada, Reno, Reno, Nevada, USA

Many studies have been conducted evaluating the use of feedback in staff training in organizational settings. Central to this literature has been the use of a variety of forms of feedback, including videotaped feedback. A distinction is outlined between video modeling and a variety of possible video feedback procedures. Previous studies have indicated a possible positive training effect on staff skills from simply being exposed to relevant videotaped performances of those skills or as a result of evaluating video modeled performances. This study evaluated the effectiveness for training teaching skills of having staff evaluate videotaped teaching skills as performed by themselves versus others. The current study results replicate earlier research on the effectiveness of evaluating video models as a training method and do not suggest a difference between observing oneself as opposed to observing someone else. Future research suggestions for the general use of video in training are provided.

KEYWORDS video modeling, staff training skills, human services

---

This article is based on a study conducted by the second author and supervised by the first author in partial fulfillment of the requirements for the Master of Arts degree. Julianne Gallinat is no longer affiliated with the University of Nevada, Reno and is now in private practice in Florida.

Address correspondence to W. Larry Williams, PhD, BCBA-D, Behavior Analysis Program, Department of Psychology, University of Nevada, Reno, 1664 N. Virginia St., Reno, NV 89557, USA. E-mail: larryw@unr.edu

Feedback has been a widely studied training and management procedure in Organizational Behavior Management (OBM) research. Although the definition of feedback has varied in published descriptions, it is generally understood to involve the process of delivering information about behavior to the person behaving (Alvero, Bucklin, & Austin, 2001; Balcazar, Hopkins, & Suarez, 1985; Ford, 1984), and that is how it will be referred to here.

Many studies have been conducted evaluating the effects of different characteristics of feedback on performance such as its frequency (Chhokar & Willing, 1984), source (Fox & Sulzer-Azaroff, 1989), mode (Emmert, 1978), and feedback in combination with other interventions (Pritchard, Jones, Roth, & Stuebing, 1988). Alvero et al. (2001) conducted a review of performance feedback as a follow-up to Balcazar et al. (1985), both of which discussed many different studies evaluating different aspects of feedback and their relative effectiveness. Results of the review conducted by Alvero et al. and supported by Balcazar et al. indicate that feedback is most effective when included in treatment packages with other procedures (e.g., feedback and goal setting).

Feedback has often involved provision of videotaped recordings of performance to those being trained on some specific task. Bricker, Morgan, and Grabowski (1972) for example, evaluated the effects of a token system along with training consisting of videotaped feedback and praise for appropriate staff behavior at a facility for children with intellectual disabilities. Results of this study suggested that videotaped feedback, praise, and the token system were effective in increasing the amount of time spent and the quality of interaction with the children. However, explanations as to the cause of the results are limited, as there was no experimental control demonstrated.

A distinction should be made here between video feedback and video modeling. Bandura (1965) and Bandura, Ross, and Ross (1963) are long-established typical examples from the learning area of Psychology of the phenomenon of modeling. In a typical example of modeling, children' behaviors subsequent to viewing a short film of other children being aggressive were associated with increases in aggression. A different example of this phenomenon of behaving similarly to a model's performance is also reported in the behavior analysis literature as imitation and generalized imitation (e.g., Baer & Sherman, 1964). Surprisingly, there has been little discussion in the behavior analytic literature of the possible similarities and differences in modeling and imitation from a behavioral analytic viewpoint. Nevertheless, it is commonly accepted that video modeling, as a training procedure, refers to the effects on a person's specific behavior after having observed a model via videotape engaging in that behavior.

Video feedback however, has been described as a training procedure in which the person being trained watches videotaped examples of their own performance with or without additional commentary or information as to the correctness or appropriateness of their performance (Panyan &

Patterson, 1974). Video feedback has also been described as a training procedure in which people observe presumably appropriate examples of targeted skills as performed by others, with the intention of increasing these performances, via modeling, in those people who watch the videotaped examples. However, video feedback can also refer to a variation in the source of feedback to a performer that involves the provision of descriptive written, spoken, or graphical information on one's performance, delivered via videotape as opposed to live interaction. A still further variation of video feedback training refers to a training procedure involving active identification by an observer, of correct and incorrect examples of someone performing a given task (typically the task to be taught) on a videotaped presentation of those performances (Dowrick & Johns, 1976).

Ford (1984) attempted to evaluate the effects of videotaped feedback. This study compared the effects of supervisor feedback, videotaped feedback, and a combination of these two types of feedback on the teaching skills of three paraprofessionals at a mental health facility. Results suggested that the largest improvement in performance was observed when supervisor and videotaped feedback were provided for a specific performance.

Panyan and Patterson (1974) evaluated the effects of only videotape feedback on the training of behavior management techniques to staff at an institution. Results suggested that videotape feedback alone had almost no effect on staff behavior.

Dowrick and Johns (1976) evaluated the effects of videotaped feedback to train more appropriate supervision of children on a psychiatric ward. Unlike the previously discussed study, in this study praise for appropriate videotaped behavior of a trainee was delivered to the trainee while they watched the videotape. One individual participated in this study and the effects of the videotape and praise were immediate and significant, suggesting that these two interventions in combination may be quite effective. However, only one individual participated, limiting the generality of the findings.

Video feedback has also been studied in combination with self-management methods. Embregts (2000) studied the effects of self-management and video feedback on inappropriate behaviors exhibited by individuals diagnosed with mild intellectual disability. In this study participants were recorded while engaging in inappropriate behaviors. Participants were then asked to watch these videos and record their own inappropriate or appropriate responses. When the participant and the person showing them the videotape correctly identified an inappropriate behavior, the participant would receive praise. During these sessions participants would also earn tokens for appropriate and inappropriate behaviors. Results suggested that use of video feedback and self-management increased participants' appropriate behaviors and decreased inappropriate behaviors. However, explanations as to the causes of the results should be considered with

caution, as praise was delivered in addition to the video feedback and self-management.

A study conducted by Harrison (2002) evaluated the effectiveness of watching a videotape of staff performing occupational safety skills. In this study, participants were asked to complete a checklist that evaluated hygiene safety-related staff performance based on a video of individuals as they performed relevant hand washing and rubber glove usage. Results suggested that the videotape scoring sessions were effective in increasing the quality of the observer's own subsequent hygiene safety performance with no other training.

Alvero and Austin (2004) reported some interesting effects on typical office safety performances by participants in an observation study. The observers' own performances improved after they conducted observations of a confederate performing the tasks while they scored them as correct or incorrect. All participants showed performance improvement as a result of watching and scoring confederate performances. In discussing their results, Alvaro and Austin (2004) suggest among other things that future studies could examine the type and specificity of the behavioral checklist as well as variation in properties of the responses observed.

It may be possible that there is differential effectiveness in observing and evaluating one's own versus other's performance. The purpose of the current study was to assess the effectiveness of having observers evaluate specific sets of teaching skill performances depicted on videotape on their own subsequent performance on those specific teaching skills. Additionally, in this study, a comparison is made between the relative effectiveness of observers judging the correctness of their own videotaped performance as opposed to other's videotaped performance.

METHOD

Participants and Setting

The present study took place at a day treatment center for adults diagnosed with intellectual disabilities. Hours of operation were from 9:00 a.m. until 3:00 p.m., Monday through Friday. Consumers represented a wide range of functioning levels and a variety of learning and behavioral issues. Fifteen consumers were involved in individualized training programs designed to strengthen or develop communication and activities of daily living as well as to develop vocational skills. These programs involved a wide variety of discrete trial assessment and teaching procedures involving modeling, shaping, fading, chaining, and reinforcement procedures, common to behavior analysis assessment and training methods.

Four female undergraduate students who were recruited to work at this organization were participants in this study. Each participant had some

exposure to the basic principles of operant conditioning via university introductory Psychology courses and some training in discrete-trial training (consisting of verbal instructions and role playing) during their initial training as new staff. Ages of the participants in this study ranged from 20–36 years of age. Participation was on a volunteer-only basis, and participants were recruited by a University Social Behavioral Institutional Review Board-approved short scripted explanation that was given to them during or shortly after their initial staff orientation period. All participants had completed brief basic introductory behavioral concepts training, including a specific session in which they were oriented on the specific teaching and intervention protocols for each of the center's consumers.

Data Collection

The dependent measure was the number of trials of acquisition training required to attain two consecutive training sessions with 100% performance accuracy. Four checklists, which evaluate the accuracy of certain aspects of discrete trial training, were used (available from the first author). Each checklist focused on a different skill area within discrete trial training. To determine the percent correct performance during a training session for a certain skill, the number of correct responses was divided by the number of incorrect and correct responses and multiplied by 100.

Skill Definitions

The skills taught to participants consisted of six specific skills from the general area labeled "preparation," five skills from the area labeled "instruction," five from the area labeled "prompt hierarchy," and six from the area labeled "reinforcement." For any training session, the participant circled either *yes*, *no*, or *n/a* on each of five trials they observed for a specific skill set. (The skill sets and definitions are available from the first author.) Examples of the skills taught for each labeled skill area and the operational definitions used to qualify them are

1. *Preparation*. Are all the work materials (training stimuli) readily available and secured by the trainer? Answer YES if the trainer has all the matching cards (there are about six of these) either in their hands or within an arms reach so a trial can be quickly started. Answer NO if the trainer does not have all the matching cards nearby.
2. *Instruction*. Was the instruction delivered while the consumer was attending to either the trainer or the work stimuli? Answer YES if the consumer was looking at the work materials (color matching cards) or at the trainer. Answer NO if the consumer was not looking at the work materials or the trainer (e.g., was looking around the room, etc.).

3. *Prompt hierarchy*. Were all of the steps delivered in order? (After the verbal instruction, gestural, and then partial physical?) Answer YES if after having given the instruction (and the consumer has incorrectly responded) the trainer removes work materials, sets out the work materials again, and delivers a gestural prompt (pointing to correct card) along with the instruction. If the consumer still responds incorrectly then the trainer would again remove work materials, re-set out the cards and then provide a partial physical prompt with the instruction. Answer NO if the trainer does not move to a gestural prompt after the instruction (Verbal prompt) and/or skips to a partial physical prompt instead of doing the gestural prompt. Answer N/A if the consumer responds correctly during this trial or if the answer to the first question is NO.
4. *Reinforcement*. Was the reinforcer delivered within 3 seconds of the response? Answer YES if the reinforcer has been given to the consumer within 3 seconds of him placing the card down on the correct match. This is easily done by counting one, one thousand, two, one thousand, three . . . after the consumer has correctly matched. Answer NO if the reinforcer is delivered to the consumer after 3 seconds of his response.

Independent Variable

There were two different conditions of the independent variable of completing a checklist while observing a person's performance. One condition, Video Model Evaluation (VME): Self, included participants using a checklist to evaluate their own performance while watching a video of themselves conducting discrete trial training trials. The other condition, VME: Other, included participants using this checklist to evaluate someone else's performance from a video of another staff conducting discrete trial training.

Experimental Design

The effects of each type of performance monitoring were evaluated through the use of a multiple baseline across skills design. The order of presentation of self versus other video content was counterbalanced for the four participants. Two of the participants were exposed to the conditions in the order ABCBC, and the other two were exposed to the conditions in the order ACBCB. (A = Baseline, B = VME: Self, C = VME: Other). All four participants were exposed to each of the four different skill checklists in the same order: preparation checklist, instruction checklist, prompting hierarchy checklist, and finally the reinforcer checklist. For example, participants 1 and 3 were exposed to the preparation checklist during the first condition,

and they were exposed to B (VME: Self). Participants 2 and 4 were first exposed to the preparation checklist during their first condition C (VME: Other). Therefore, participants 1 and 3 were exposed to the preparation and prompting hierarchy checklists while evaluating videos of themselves, whereas participants 2 and 4 were exposed to those checklists while evaluating others' performance, and in the opposite order for the other two skill set checklists.

Training

Prior to evaluating any videos during the experimental conditions all participants were trained on how to use the checklists through explanations by research assistants. These explanations consisted of how to fill out the individual checklists and answer the questions. Research assistants taught these checklists using the predetermined operational performance criteria (available from the first author) to ensure that explanations would not vary.

With respect to the actual discrete trial skills, all participants were trained during two 1-hour sessions with a trained staff member as to how to carry out these discrete trials. Training consisted of reading through the protocol and modeling (e.g., the participant watched a staff member carry it out). All participants experienced the same general teaching program content throughout the training sessions. A color matching training session with the same consumer was used during each training session in every condition. The only exceptions were the generalization probes, which were conducted after the study and evaluated participants while they were conducting a different type of training session with the same consumer.

Baseline

During this phase participants conducted training sessions as usual. Participants were videotaped while conducting training sessions with the consumer Teri, at the day center. Teri was a 43-year-old male diagnosed with severe intellectual disability who was at the center receiving vocational, communication, and daily living skills training. A research assistant collected data by filling out the performance checklists (for each of the four skill areas) each time the participants conducted a specific training session with the target consumer. The five trials chosen to be evaluated by the research assistants were arbitrarily chosen by the research assistant and did not follow any specific pattern. The trials that were evaluated were any consecutive combination of trials, such as 1 through 5, 2 through 6, 6 through 10, or 16 through 20. Research assistants decided which trials to evaluate by looking back at previous sessions and making sure that they did

not evaluate the same sequence of trials more than twice in a row. Therefore data reflected the beginning, middle, and end of the twenty trial sessions. For all participants there was a wide range of the relative position of trials evaluated across the possible 20 trials in any given session.

VME: Self

During this phase participants watched a video of themselves conducting a training session while completing the performance checklist for one of the four skill areas. Participants evaluated themselves conducting 5 of the 20 trials that made up a training session. The trials being evaluated by participants were chosen using the same procedure described above for the research assistants' evaluations of participant training performance. In addition, the session chosen to be observed was different each time and used videos from the most recent sessions conducted by participants.

Once participants completed the performance checklist for one skill, research assistants would tell participants how they scored themselves (by calculating a percentage correct from the checklist). Participants were shown a video of themselves during each consecutive training session (of consumer color matching) they conducted during this phase. Once participants met the criteria (as determined by a research assistant's scores) for two consecutive sessions of 100% on the performance skill checklist for one skill area, they were exposed to the next skill checklist and moved on to the evaluation of videotaped performance of someone else. Data continued to be collected by a trained research assistant on the performance of participants while conducting relevant discrete trial training sessions in all of the 4 skill areas.Each participant was exposed to a maximum of four video-modeling sessions a week. Duration of the video evaluation sessions was approximately 5 to 10 minutes.

VME: Other

This phase was conducted exactly as the VME: Self phase except that the participant was exposed to videos of others conducting training sessions and used the relevant checklist to evaluate someone else's performance on a block of five trials for a specific skill set.

Generalization Probes

For three of the participants at least two generalizations probes were conducted with the participant running a different type of training session with the target consumer. The training session labeled transitions included

having the consumer move around the room while matching colored cards presented by a trainer. This was similar to the color matching program in that the consumer was matching colors, but different in that the training session was no longer taking place at a table while sitting down. Maintenance probes were also conducted for each participant after completion of the final training phase by observing the participant conduct the color matching training session.

Interobserver Agreement

Interobserver agreement (IOA) was conducted for 57% of all trials by trained research assistants independently watching the videotapes of the participants conducting the training sessions and filling out the performance checklists. Each research assistant was trained by watching practice videotapes that depicted other research assistants and the second author performing discrete-trial training sessions and scoring the performance. Feedback was delivered to the assistants on accuracies and inaccuracies. Research assistants were considered trained when they had correctly scored two consecutive training sessions at or above 90% accuracy. IOA was calculated by using an exact trial by trial agreement measure. For a block of 10 trials, for example, on each trial, each question was considered a match or a non match. The matches and nonmatches were calculated across all checklists. Then, all the matches were divided by the total of matches and nonmatches, and multiplied by 100.

The IOA data are shown in Table 1. As can be seen, the mean IOA for each participant for the skill sets and conditions overall was 95% (range 89% to 100%). The IOA scores for each condition for each participant are stable, indicating that one checklist was no more subjective than the other. However, IOA collected during the follow-up phases, which included generalization probes as well as maintenance probes, was lower than the IOA during any other condition. This could be attributed to the fact that a different program was being evaluated, and the questions had to be adjusted to the new program. In addition, during the experimental phases most disagreements between research assistants scoring the performance using the checklists were found on the instruction skill checklist as well as on the reinforcer checklist. On both of these checklists there were similar questions attempting to evaluate the level of intonation of the participant's voice, a fairly subjective measure. Data, which were collected on the accuracy of participant responding during the video feedback conditions, indicated that participants were responding with 95% accuracy across all four checklists.

TABLE 1 Interobserver Agreement Measures for All Participants Across All Conditions

Conditions	Participant 1	Participant 2	Participant 3	Participant 4	All participants
Number of IOA Sessions	12	8	10	10	40
Percentage of All Sessions	50%	44%	67%	67%	57%
Average IOA	95%	96%	93%	95%	95%
Baseline IOA	99%	100%	99%	100%	99%
Preparation Skills	97%	98%	97%	99%	98%
Instruction Skills	96%	91%	87%	97%	93%
Prompt Hierarchy Skills	95%	98%	94%	97%	96%
Reinforcer Skills	93%	98%	95%	95%	95%
Follow-up	93%	93%	92%	90%	92%

Note. Table 1 shows the number and percent (rows 1 and 2) of all sessions for which IOA measures were taken for each participant (columns 1–4) and overall (column 5), as well as the average IOA scores for each participant across all sessions (row 3), which were then calculated for baseline and each intervention phase (rows 4–9).

Preference Survey

A survey was given to the participants a week or two after they completed all four phases of the study. The survey asked three questions for which the participants were asked to circle their answer. The questions asked were

1. Which do you prefer to do, observe yourself or someone else?
2. If you had a choice between watching a video and scoring it or receiving directverbal feedback from your supervisor, which would you choose?
3. During which condition, observing yourself or someone else, did you feel as though you could better understand what you were supposed to do during discrete trial training?

RESULTS

The multiple baseline across skills strongly suggests that performance was improved as a result of the introduction of the video evaluation procedure (see Figures 1–4). As one can see, for most of the participants performance improves to 100% correct responding after the introduction of even only one video evaluation session. The use of a counterbalanced order of video model type across participants also suggests the performance improved as a result of the first evaluation method employed (either self-as-model or other-as-model intervention) rather than as the result of the type of model.

The number of trials to criterion during each phase did not seem to indicate that any one condition (self- versus other-as-model) was more effective at improving performance than another (see Figures 1–4).

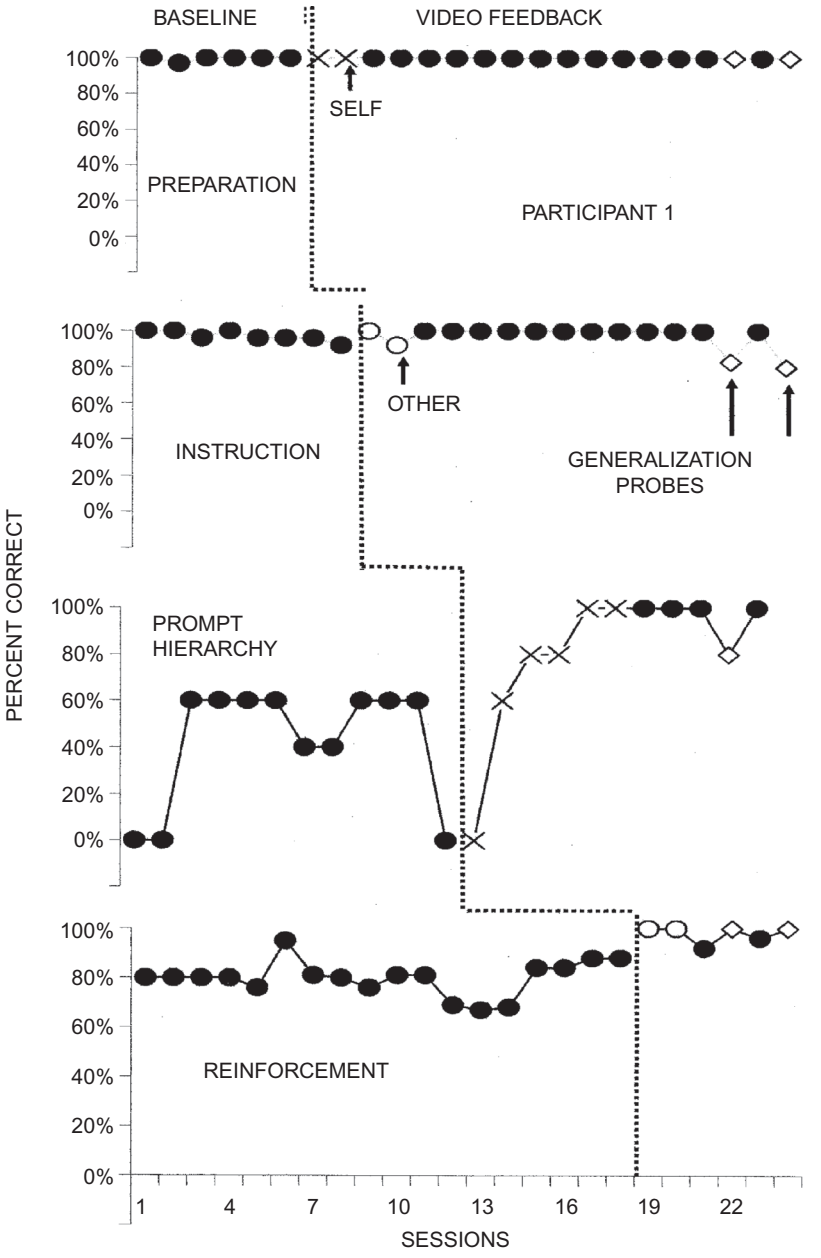

FIGURE 1 Percent correct responding for participant 1 before and after video model evaluations in a sequential fashion across the four skills of session preparation, instructions, prompt hierarchy, and reinforcement.

In addition to the number of trials to criterion, data were also analyzed with respect to the percent correct gain per skill, per participant, for video model self versus other (Figure 5). Averages of the two data sets, self versus other, are almost identical, again indicating little differentiation

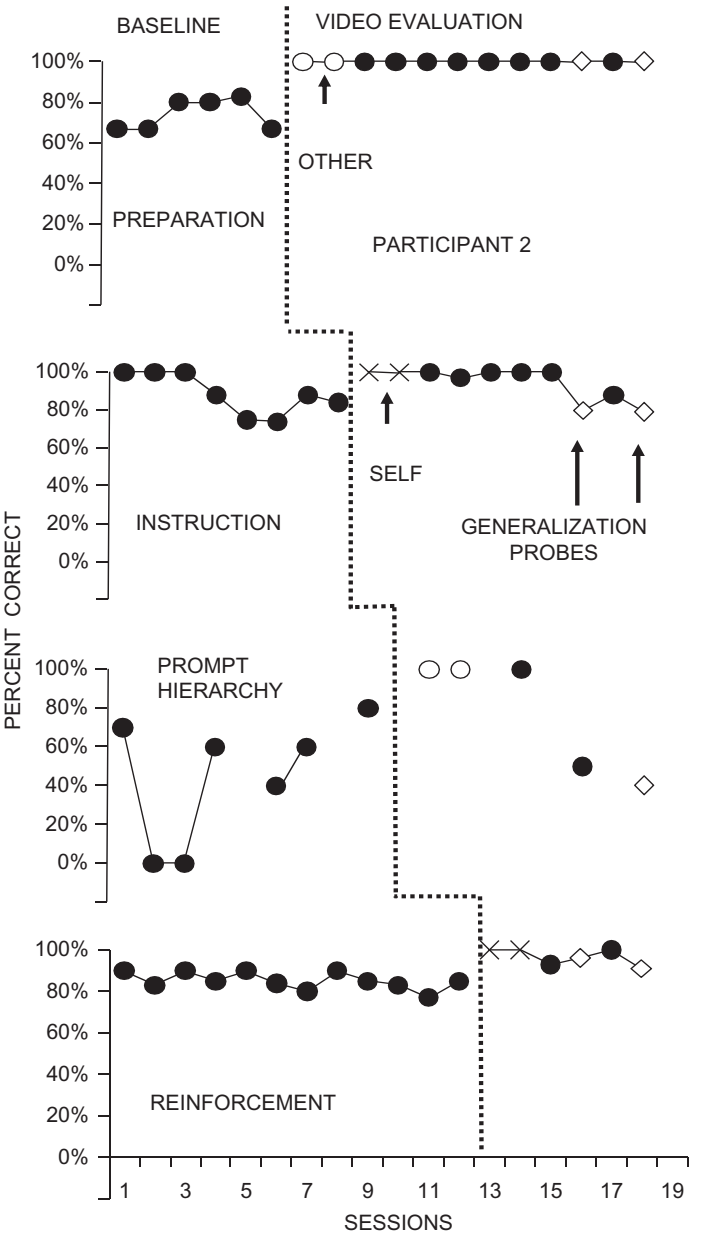

FIGURE 2 Percent correct responding for participant 2 before and after video model evaluations in a sequential fashion across the four skills of session preparation, instructions, prompt hierarchy, and reinforcement.

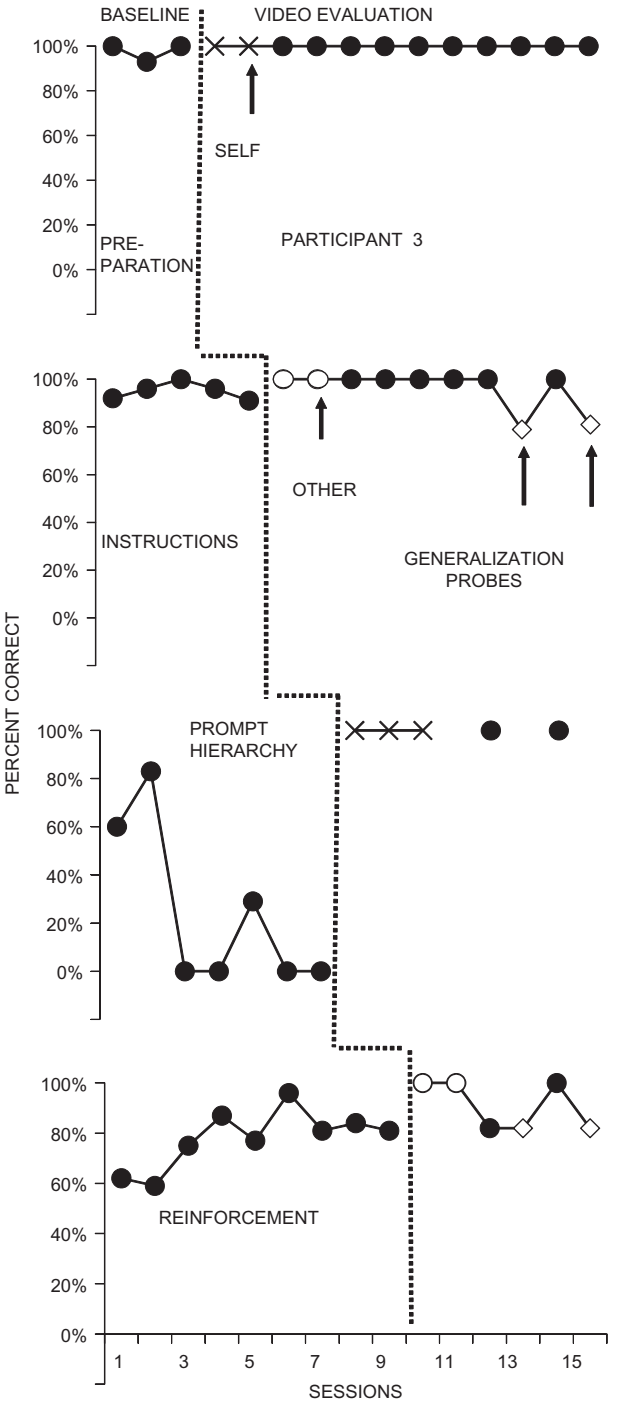

FIGURE 3 Percent correct responding for participant 3 before and after video model evaluations in a sequential fashion across the four skills of session preparation, instructions, prompt hierarchy, and reinforcement.

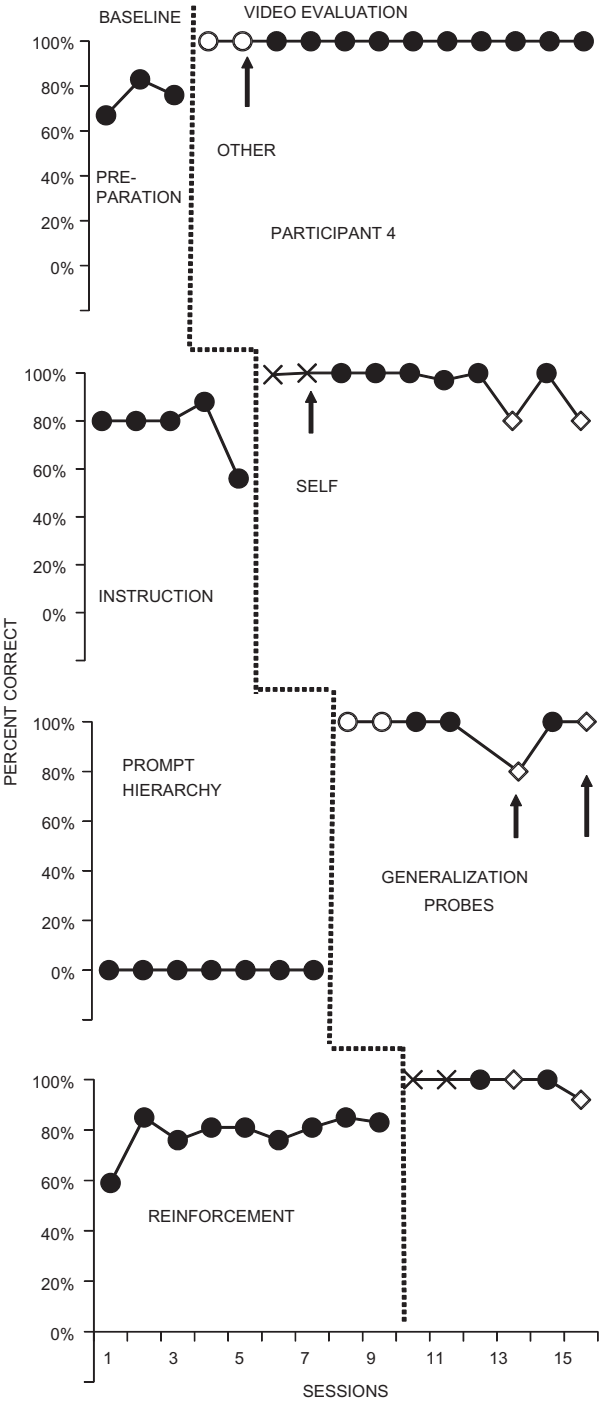

FIGURE 4 Percent correct responding for participant 4 before and after video model evaluations in a sequential fashion across the four skills of session preparation, instructions, prompt hierarchy, and reinforcement.

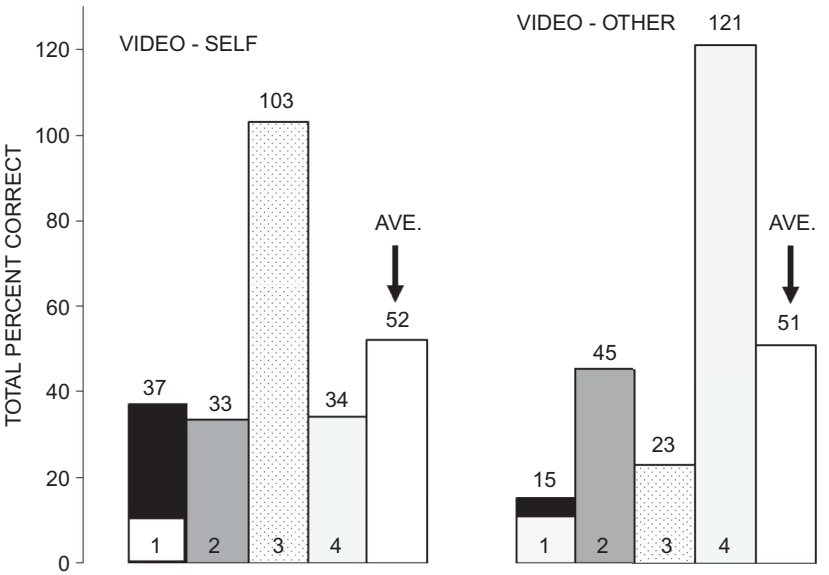

FIGURE 5 Total percent correct gain per skill, per participant, from baseline to intervention.

between video evaluation of self versus other. The individual participant acquisition data (Figures 1–4) and these data also suggest that the preparation, instruction, and reinforcer skills checklists were not as sensitive and had ceiling effects for participants 1 and 3. On the other hand, participants 2 and 4 seemed to display less ceiling effects for these skill checklists.

Figure 6 provides a summary description of the actual accuracy each participant showed during the video evaluation sessions. All participants evaluated the training videos for preparation skills at 100% accuracy. Participant 1 scored 100%, 95%, and 87% accuracy for instruction, prompt hierarchy, and reinforcement skills respectively. Participant 2 scored 95%, 100%, and 90% respectively for these same skills. Participant 3 scored 87%, 78%, and 90% respectively for the same skill sets. Finally, participant 4 scored 100%, 100%, and 95% accuracy, respectively for the skills of instruction, prompt hierarchy, and reinforcement.

Although the training data for the consumer who was being trained during the color matching training session in this study indicated improvement, permission was not available to present these data.

Results of the survey given to participants upon completion of the study suggested that whereas participants preferred to watch a video of someone else, they better understood how to conduct discrete-trial training programs after watching and evaluating themselves. Survey results also suggested that participants preferred video modeling evaluation to supervisor feedback.

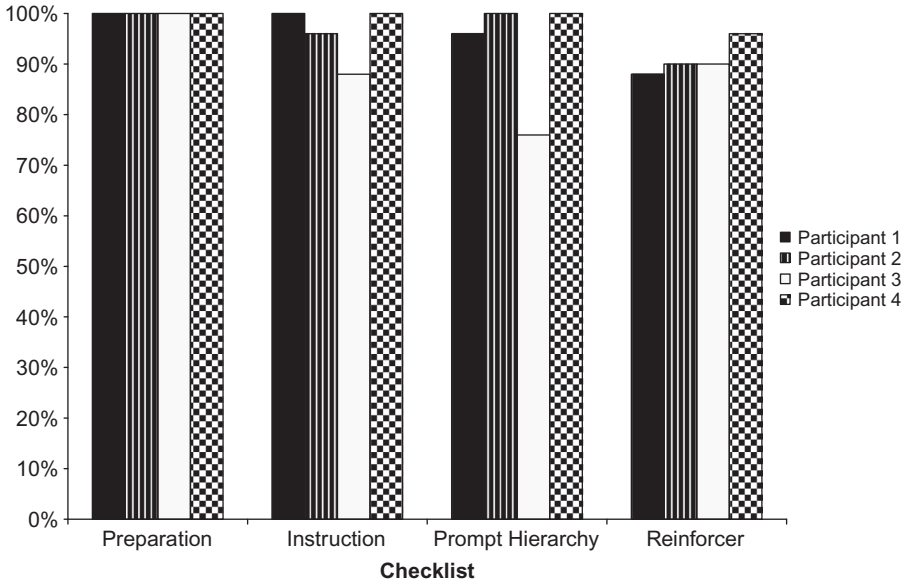

FIGURE 6 Accuracy of participant's behavior of scoring performance using the checklists during the video model evaluation phases.

DISCUSSION

Results of the present study support the findings of Alvero and Austin (2004), Ford (1984), and Harrison (2002), suggesting that videotape feedback involving scoring of relevant performance viewed via video recording is an effective training tool for improving staff performance. The data demonstrated immediate and significant increases in performance upon introduction of video evaluations with checklists regardless of the person in the video examples, suggesting that both forms of video models were equally effective.

The follow-up data also suggest that performance maintained for a period of time (two weeks) after participants were exposed to the video evaluation sessions. Maintenance probes indicated that participants' performance maintained well above baseline. Data collected during generalization probes indicated that performance was slightly less correct than during the maintenance probes. One may assume that this is a result of the effects of video feedback not generalizing to other programs. However, the program observed to assess generalization was somewhat different than the color matching program that was used during the study. Many aspects of these two programs, from the instructions delivered to the protocol for carrying out the prompting hierarchy, were different. In addition, although performance data are lower during the generalization probes, they are higher than the initial baseline phases.

The results did not seem to suggest that effectiveness differs with respect to evaluating performance from a video model of oneself versus another individual. Participants 2, 3, and 4 all required the same number of trials during each phase to reach the criterion of 100% or better for two consecutive sessions. It took each of the previously mentioned participants only two sessions to meet this criterion. The exception was participant 1, who took a total of eight sessions to meet the criterion during the VME: Self condition, while it only took this participant six total sessions to meet the criterion for the VME: Other condition. However, the small difference in the number of trials between the two phases for this participant is not necessarily suggestive that watching someone else is more effective than watching oneself. Rather, it suggests that this participant had a difficult time with one of the checklists to which they were introduced. Data, which displayed the percentage correct gain per skill per participant, also did not indicate a difference between the two conditions when an average measure was calculated. In conclusion, the data not displaying differentiation between phases for any of the four participants supports a position that neither type of video feedback was more effective than the other.

This conclusion is also supported from an analysis of the role of accuracy in the participants' evaluation accuracy during video evaluation training. Comparing the procedure associated with most gain to the skill set associated with the least accuracy for each participant, we see that participant 1 responded better on VME: Self evaluation (37% gain vs. 15% gain), but that gain was associated with skill set 3, which had the lowest accuracy. Participant 2 responded better to VME: Other evaluation (45% gain vs. 33% gain), but that gain was for the lowest accuracy skill sets (instruction and reinforcement). Participant 3 had much better improvement associated with VME: Self (103% vs. 23%), but the large gain was for the least accurate evaluations (instruction and prompt hierarchy). Participant 4 responded better to VME: Other (121% vs. 34%), and that gain was associated with a higher accuracy (prompt hierarchy vs. reinforcement). Thus, the better performance for participants 1 and 3 on VME: Self are associated with skill sets for which they were the least accurate in training. Participant 2 also showed the most gain for the skill sets for which they were least accurate. This analysis indicates that the best response for three of four participants was to training on skill sets for which they were least accurate at evaluating, rather than if the training involved evaluation of models that were themselves as opposed to other staff.

A limitation of the present study is that the overall data do seem to indicate somewhat of a "ceiling effect." First of all, for three of the four participants their baseline preparation and instruction skills were already at 90% or better before even introducing the checklists. After introducing the checklist, each of these participants' skills for these two areas, preparation and instruction, did maintain at 100% correct, but the intervention did not

allow for a more sensitive measure of whether or not one may have done a better job. Perhaps had the skills been more difficult to carry out, the current arrangement may have resulted in a differentiation of video model type as an independent variable.

One may also note that results did not support the findings from Panyan and Patterson (1974), which suggested that video models without the combination of someone else providing feedback is ineffective. During the present study no additional feedback regarding performance was delivered, only feedback generated by the participants themselves. Feedback regarding accuracies or inaccuracies was never delivered while the participants were filling out the performance checklists. Research assistants did calculate and communicate a percentage correct score after participants had conducted the evaluations on the checklists. However, the score delivered was based on the participants' recorded evaluations of their own or others' performances. Results from this study suggest that further studies should evaluate whether there is a need for supervisor feedback in combination with participant generated evaluations of videotaped performance of themselves or others. For instance, it may be that feedback delivered on the accuracy of performance assessed for participant evaluation of video models may result in even larger increases in actual teaching performance and accuracy. In this study there was no feedback delivered as to the accuracy of the participant's scoring, in an attempt to isolate the effectiveness of the scoring of video models alone; however, even quicker or stronger increases in behavior may have been observed if accuracy feedback was delivered to participants at the time of their scoring.

Another possible limitation of this study was the potential confounding variable of explaining the checklist to the participant prior to the video evaluation phases. The scripts, which were used to explain how to fill out the performance checklists, could be considered an instructional prompt that was experienced prior to the video modeling phases. However, participants were given these instructions as well as exposed to live role modeling prior to any baseline data being collected. Therefore, one would expect that participants' baseline performances would reflect the effectiveness of these instructions. The data clearly indicate that performance markedly improved and maintained for all of the skills upon introduction of evaluating videotaped example performances.

As previously mentioned, the results do not suggest that one type of video model was more effective than the other. However, it may be that a "ceiling effect" present in this study did not allow for a sensitive comparison to be made. This observed baseline ceiling effect may have been the result of the initial checklist training. Future research may limit such training or attempt to develop a more difficult or sensitive measure to attempt to see if there is a difference between the two forms of video presentation. For example, one may wish to evaluate whether there is any

difference in watching videotapes of one's own versus others' performance on contrived predetermined teaching skill exemplars as compared to monitoring ones' own or others' performance while actually conducting training sessions.

Additionally, it may be interesting to further study other characteristics possibly responsible for the effectiveness of several types of videotape usage. For example, it may be that videotaped model performance evaluation or other types of video format feedback on previous performance, delivered shortly before (e.g., one week, day, hour) the time in which the person must perform again is most effective. In this study, participants evaluated a video prior to each session. It may have been that if they watched or evaluated a video a day or a week beforehand, then their performance may have been affected differently. Similarly, there may be some combinations of variables associated with videotape feedback that may be most effective.

Future research also may wish to evaluate whether or not the information delivered via video was necessary. It may have been that participants could have observed others or themselves and scored performance "in real time," and we would have still seen the same improvement in their behavior. It may be interesting to investigate if there are any differential effects of monitoring through different modalities.

REFERENCES

Alvero, A. M., & Austin, J. (2004). The effect of conducting behavioral observations on the behavior of the observer. *Journal of Applied Behavior Analysis*, *37*, 457–468.

Alvero, A. M., Bucklin, B. R., & Austin, J. (2001). An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings. *Journal of Organizational Behavior Management*, *21*(1), 3–29.

Baer, D., & Sherman, G. (1964). Reinforcement control of generalized imitation in young children. *Journal of the Experimental Analysis of Behavior*, *1*, 37–49.

Balcazar, F., Hopkins, B. L., & Suarez, Y. (1985). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management*, *7*(3/4), 65–85.

Bandura, A. (1965). Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of Personality and Social Psychology*, *1*, 589–595.

Bandura, A., Ross, D., & Ross, S. A. (1963). Imitation of film-mediated aggressive models. *Journal of Abnormal and Social Psychology*, *66*, 3–11.

Bricker, W. A., Morgan, D. G., & Grabowski J. G. (1972). Development and maintenance of a behavior modification repertoire of cottage attendants through T.V. feedback. *American Journal of Mental Deficiency*, *77*(2), 128–136.

Chhokar, J. S., & Walling, J. A. (1984). A field study of the effect of feedback frequency on performance. *Journal of Applied Psychology*, *69*(3), 524–530.

Dowrick, P. W., & Johns, E. M. (1976). Video feedback effects on therapist attention to on-task behaviors of disturbed children. *Journal of Behavior Therapy and Experimental Psychiatry*, *7*, 255–257.

Embregts, P. J. (2000). Effectiveness of video feedback and self-management on inappropriate social behavior of youth with mild mental retardation. *Research in Developmental Disabilities*, *21*, 409–423.

Emmert, G. D. (1978). Measuring the impact of group performance: Feedback versus individual performance feedback versus individual performance feedback in an industrial setting. *Journal of Organizational Behavior Management*, *1*(2), 134–141.

Ford, J. E. (1984). A comparison of three feedback procedures for improving teaching skills. *Journal of Organizational Behavior Management*, *6*(1), 65–77.

Fox, C. J., & Sulzer-Azaroff, B. (1989). The effectiveness of two different sources of feedback on staff teaching of fire evacuation skills. *Journal of Organizational Behavior Management*, *10*(2), 19–77.

Harrison, C. (2002). *The effects of trained observation on the hygienic behaviors of the observer* (Unpublished master's thesis). University of Nevada, Reno.

Panyan, M., & Patterson, E. T. (1974). Teaching attendants the applied aspects of behavior modification. *Mental Retardation*, *12*, 30–32.

Pritchard, R. D., Jones, S. D., Roth, P., & Stuebing, K. K. (1988). Effects of group feedback, goal setting, and incentives on organizational productivity. *Journal of Applied Psychology*, *73*(2), 337–358.