EFFECTS OF MULTIPLE EXEMPLAR TRAINING ON GENERALIZATION OF PREVOCATIONAL SKILLS

by

Blane L. Brown
B.S.E., Southwestern Oklahoma University, 1979

Submitted to the Department of
Education and the Faculty of the
Graduate School of the University
of Kansas in partial fulfillment
of the requirements for the degree
of Master of Science

Redacted Signature

Professor in Charge

Redacted Signature

Redacted Signature

Committee Members

Redacted Signature

For the Department

Date thesis Accepted

ABSTRACT

This study was designed to increase generalization of severely multiply handicapped students. The intent was to determine the effect of generalization on trained and untrained vocational items. Two severely multiply handicapped students, ages 17 and 18, participated in the research. A multiple baseline design was used to illustrate the rate of acquisition of the trained and untrained items in multiple exemplar and nontrained items. Results of this study indicate that the use of a multiple exemplar item can be an effective way of increasing the rate of generalization on vocational tasks for some severely multiply handicapped students.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

The ability to apply what has previously been learned to new situations has been termed generalization. Generalization appears to be a natural process in most nonhandicapped persons, but several investigators have documented that students with severely handicapping conditions do not exhibit the skills they have learned across settings, persons, or objects without specific generalization training (Stokes & Baer, 1977; Wehman, Abramson, & Norman, 1977; Brown & York, 1974).

Stokes and Baer (1977) discuss nine procedures for promoting generalization starting with the two most frequent and the least analytical procedures: train and hope and sequential modification. The first of these two procedures is used when the potential for generalization is recognized, the presence or absence of generalization is noted, but no particular effort is expended to accomplish generalization. The second procedure, sequential modification, uses procedures to effect changes in nongeneralized conditions. Although contributing significantly to our understanding of the generalization of behavior change programs, these concepts are not examples of the programming of generalization. The next seven procedures discussed by Stokes and Baer (1977) begin with the use of natural contingencies. According to this procedure, generalization can be programmed by introducing natural reinforcement contingencies which refine and maintain skills without further therapeutic intervention.

The training of sufficient exemplars is numerically the most extensive area of programming. Generalization to untrained stimulus conditions and to untrained responses is programmed by the training of sufficient exemplars of those stimulus conditions or responses. Train loosely is a programming technique in which training is conducted with relatively little control over the stimuli and responses involved, and generalization is thereby enhanced. In the procedure of indiscriminable contingencies, the contingencies of reinforcement or punishment, or the setting events marking the presence or absence of those contingencies are made less predictable, so that it becomes difficult to discriminate reinforcement occasion from nonreinforcement occasions. Common stimuli may be employed in generalization programming by incorporating it into training settings that are salient in generalized settings, and that can assume functional or obvious roles in the training setting. Mediated generalization requires establishing a response as part of new learning that is likely to be utilized in other problems as well, and thus result in generalization. The final procedure, train to generalize, involves reinforcing generalization itself as if it were an explicit behavior.

Wehman, Abramson, & Norman (1977) did a selective review of behavior modification programs that emphasized procedures for programming transfer of training and response maintenance. Their principle focus was to evaluate the effectiveness of generalization techniques utilized with exceptional children and to identify several points for maintaining performance in the classroom. These investigators found that the most substantial factors in programming transfer of

training and response maintenance were procedures to vary stimulus conditions, parent training, and peer programming.

Drabman, Hammer, & Rosenbaum (1979) provide a conceptual framework with various generalization effects of behavior modification programs. A large number of child behavior modification studies were reviewed and categorized according to guidelines of a generalization map. Suggestions are offered on methodology and design for the future research.

The Individualized Curriculum Sequencing Model outlined by Holvoet, Mulligan, Schussler, Lacey, & Guess (1982) proposed that teachers should use varied materials, either within or across sessions, in order to maintain student interest and to enhance generalization. This recommendation is supported by the findings of Horner & McDonald (1982) which demonstrated that general case instruction resulted in better generalization than did single instance training with severely handicapped students. These investigators used three different varieties of capacitors in general case training format, and only one variety of capacitor in the single instance training. They then tested generalization by having the student work with novel, untrained types of capacitors.

The present investigation is similar to the study by Horner & McDonald (1982), in that a vocational task was studied. It differs, however, in that the various exemplars (and probe items) all require slight variations in the response and were quite dissimilar in appearance. This study also does not attempt to compare multiple exemplar to single instance training, but seeks only to establish whether multiple exemplar training that requires variations in a

response results in generalization to untrained items requiring similar responses.

Schworm and Abelseth (1978) noted that it is necessary for individuals to make generalization across materials with similar properties. The teacher may find that the severely multiply handicapped student may become attached to familiar training materials and find it difficult to use other devices with similar properties. By presenting the student with materials with similar properties during acquisition of a response, generalization may be promoted.

The purpose of education is not only to teach skills but to furnish students with a consistent introduction to events and activities which increase their independence from non-normal support systems, increase integration with non-handicapped individuals, and facilitate participation in the community (Brown, Nietupski, & Hamre-Nietupski, 1976). To achieve these objectives, Horner & McDonald (1982) suggest the following two ways: a) students need to learn skills that function, not only in specific teaching situations, but in nontrained natural situations as well; and b) teachers need a generalized performance across nontrained settings, nontrained people, new materials, and situations in which trainers are absent.

CHAPTER II

METHOD

Subjects

A male and female student were selected from a public school serving adolescents with special needs. Both students were selected from a classroom for severely handicapped individuals. The two students were Arthur and Kim, 17 and 18 years of age, respectively. They were classified as severely retarded using the MS/CA ratio on tests such as WISC-R as well as on adaptive behavior measures such as the TARC Assessment Guide (Sailor & Mix, 1975).

Arthur could be described as a person in constant motion who stared off into space and flicked his fingers when not attending to the task given to him. Otherwise, he had good fine motor skills and enjoyed working on the tasks he was given.

Kim could be described as one who liked learning the tasks and knowing how to assemble them. She did not require many verbal, demonstration, or physical cues, and seemed to learn quickly.

Setting

The study took place in a large, well-lighted, vocational classroom of severely multiply handicapped students which was served by staff from a federally funded demonstration project. Each student was worked with individually. The student was seated at a long flat work bench next to the investigator. The reliability observers sat about five feet away, to the right of the student. Five other students and three staff were nearby working on leisure, vocational, and academic tasks unrelated to

the research.

Tasks/Materials

There were six tasks used in this study. Three tasks served as the training items: assembling a metal faucet; screwing a metal hinge into a piece of wood; and screwing a metal handle into a piece of wood. Generalization probe items consisted of screwing a large bracket into a piece of wood, assembling a deadbolt lock on a piece of wood, and screwing two wooden handles into a piece of wood.

Training items. Each of the training tasks was divided into several sub-tasks (steps) that were taught in each session. On the metal faucet task, the student was required to place a knob on the faucet and then place one screw in a hole on top of the handle. The correct screwdriver was then chosen, and the screw tightened.

On the metal handle task, the student was required to place the handle in predrilled holes in the board, put the screw in by hand on the underside of the board, select the correct screwdriver, and then tighten the screws.

The metal hinge task required the student to place the hinge on a board, select the correct screwdriver, and place and tighten four screws.

Generalization items. Each of the generalization probe items was divided into several subtasks that were analyzed in each session. On the large bracket, the subject was to pick up the bracket and screw, place the bracket over the hole in the board, place the screw in the hole, select the correct screwdriver, and tighten the screw.

The wood handle required that the subject first place two screws in

predrilled holes, turn the wood over, and tighten the wood handles by hand. The student next needed to select the correct screwdriver, turn the board over again, and tighten the two screws.

On the deadbolt lock, the student was required to place four screws in the lock holes, select the correct screwdriver, tighten those four screws, place the clasp portion on the board, put in two more screws, and tighten them.

Measurement

The dependent variable in this study was measured on each of the six tasks. This was done for the training tasks by recording the level of prompting needed for a student to complete each step of a task correctly, using the least-prompts strategy described by Lent and McLean (1976). This strategy specifies four levels of prompting which are presented in sequence until the student makes an acceptable response. The levels consisted of: (+) = the subject performed independently; (V) = a verbal prompt is given; (D) = a demonstration prompt is given; (P) = the student is given physical assistance in completing the task. At the end of each task the number of steps the student had completed correctly at the independent level was determined. This total was then converted into a percent by dividing that total by the number of steps in the task.

In the generalization probe tasks, the investigator recorded whether or not the student completed any or all of the steps of the task independently. The total number of independent responses was then dividided by the number of steps in that task and then converted to a percentage.

Procedures

Baseline condition. In the baseline condition, the student sat next to the experimenter at the long work bench. The instructor gave the student the materials for one of the six tasks and three different types of screwdrivers were placed in front of the student. The order in which different tasks were given to the student were prearranged, and differed in consecutive sessions. The student then had an alloted amount of time to complete the task. For example, the student would be given the materials for the metal hinge and if he/she began working, he/she was given ninety seconds to complete the task. If the student did not pick up the materials and begin working in the first thirty seconds, then materials were removed and the next task was presented. If the student finished the item or quit working, the materials were removed after thirty seconds of "no work", or at the end of the ninety seconds, whichever occurred first. The responses were scored as correct (+), no response or incorrect response (-), and approximation (A). The instructor praised the student after the task was over by saying "that was a nice job" or "you worked pretty good on the (item)". Then the instructor would give the student a different item to assemble.

Multiple exemplar condition. In the multiple exemplar condition, the student was again seated next to the experimenter at the work bench. The instructor gave the student either a generalization probe item or a training item to assemble. When the student was given a training item, the instructor would say, "put this together". If the student did not respond correctly, the instructor would give a verbal, demonstration, or a physical prompt on the steps of the task. The responses were recorded

on the data sheet as (+) for independent, (V) for verbal, (D) for demonstration, and (P) for physical prompt. Praise was provided specific to the step of the task the subject was engaged in. Arthur also received a pat on the back, but Kim did not. When the student was given a generalization probe item, the instructor would hand the item to the student and say "put this together". The student's responses were recorded as correct or incorrect on each step of the task. The student was given the same amount of time to complete the probe items as in the baseline condition. The training and probe items were given in a different order every session. Praise was given at the end of the task, e.g., "nice job completing the (item)" or "I liked the way you worked on the item".

Experimental Design

Each task was analyzed in a multiple baseline format across subjects (Kazdin, 1973) to evaluate the effectiveness of using multiple exemplar training to enhance generalization. All tasks, whether training tasks or generalization tasks, were assessed for several sessions to determine the student's base level of performance without training. Multiple exemplar training was then initiated across all three training tasks and the effect was assessed by observing the student's performance on the three generalization probe tasks. The onset of training was delayed more for Kim than it was for Arthur in this design to increase internal validity.

Reliability

Reliability of the data was assessed for each student several times during the baseline and multiple exemplar conditions. Reliability

consisted of the instructor and an observer concurrently and independently recording the level of prompting on each step of the trained task, and the correctness of the response(s) on each step of the probe tasks.

The instructor sat by the student at the work bench while the observer sat five feet away on another smaller work bench. At the end of the session, the instructor compared the two data sheets to see how many agreements there were between the two observers.

An agreement was scored if both the instructor and observer recorded the level of prompting in the same way on a step (e.g., both experimenter and observer put a (V) to indicate that a verbal prompt was needed to select the correct screwdriver on a task. Reliability was calculated by dividing the number of agreements by the number of agreements and disagreements, then multiplying by 100.

CHAPTER III

RESULTS

Reliability

The reliability measures for Arthur in baseline on the trained items were 90% on the metal faucet, and 100% on both the metal hinge and metal handle. For the generalization probe items in this condition, the reliability measure were 100% on all three items. The reliability measures in the multiple exemplar condition for Arthur on the trained and generalization probe items were 100% on all items.

The reliability measures for Kim in the baseline condition for the trained items were 100%. Reliability measures in this condition, for the generalization probe items were 93% with a range of 86%-100% for the wood handle, with 100% on both the large bracket and deadbolt lock. The reliability measures in the multiple exemplar condition for the trained items were 100%. The reliability measures in this condition for the generalization probe items were 86% on the large bracket with a range of 75%-100%, and 100% for both wood handles and the deadbolt lock.

Training Items

Figure 1 through 3 include the data from the three training items for Arthur and Kim respectively. Figures 1 through 3 illustrate, graphically, the percent of correct responses made during each training session on the three training items (metal faucet, metal handle, and metal hinge). The training items are presented with Arthur's data on the top graph and Kim's on the bottom graph. These data are presented in this ways for both baseline and multiple exemplar conditions. These

tasks were directly taught to the students during the multiple exemplar condition.

Item 1 (metal faucet). The baseline data for Arthur shows a decending trend with a mean score of 27%. In the multiple exemplar condition Arthur shows a generally accelerating trend. His mean score during training was 83%. Kim's data exhibit some acquisition during the third session of the baseline condition. The third session had 40% correct response and stayed at that level throughout baseline. Kim's baseline mean was 31%. Kim's data during multiple exemplar training showed a clearly ascending trend, and reached criterion on the eleventh session. Her mean score was 75%.

Item 2 (metal handle). The baseline data for Arthur revealed no acquisition on the metal handle (mean score was 0%). The multiple exemplar condition data for Arthur had an ascending trend with a mean score of 88%. Kim showed no acquisition in the baseline condition which gave her a mean score of 0%. Kim's data in the multiple exemplar condition also showed an ascending trend with criterion reached on the second session. Her mean score was 94%.

Item 3 (metal hinge). Arthur's baseline data initially showed 8% correct responses, then descended to zero for the next two sessions. Arthur's mean score in this condition was 3%. The multiple exemplar condition data on this item show that he reached criterion on the third training session. His mean score in this session was 88%. Kim showed a zero performance on the metal hinge during the baseline condition for a mean score of 0%, over the eight sessions. The data for Kim in the multiple exemplar condition showed a quick ascending trend, reaching

criterion on the second training session. Her mean score in the multiple exemplar condition was 91%.

Generalization Probe Items

Figures 4 through 6 illustrate, graphically, the percent correct responses made during each generalization probe session on the three probe items (large bracket, wood handle, and deadbolt lock). The generalization probe items are presented with Arthur's data on the upper graph and Kim's on the lower graph. These data are presented in this way in both conditions. No direct training occurred during the multiple exemplar condition for these items.

Item 4 (large bracket). During baseline, Arthur showed one day in which he made one correct response out of four possible. The other two sessions were at zero. His mean score was 8%. Arthur's multiple exemplar data showed that he quickly showed generalization. Arthur's mean score was 92% in this condition. Kim showed no generalization during the eight sessions of baseline. She then exhibited a gradual learning of the task without training in the multiple exemplar condition. She had a mean score of 51%, but reached criteria and remained there on session eleven.

Item 5 (wood handles). Arthur's baseline data showed no generalization. Once he was in the multiple exemplar condition, he showed a rapid and high degree of generalization. He reached criterion four sessions in a row, had one session of poor performance, but returned to criterion on the next session. Arthur's mean generalization score was 93%. Kim's data show that after one day she acquired one correct response out of seven. That was maintained all through baseline condition. Her mean score was 13%. During this multiple exemplar

condition, Kim showed an immediate increase in performance when multiple exemplar training was implemented, followed by a variable, but generally high, performance. Her mean performance on this item was 94%.

Item 6 (deadbolt lock). The baseline data displayed zero generalization for Arthur. Generalization performance of 100% on this item was reached in the second session of the multiple exemplar condition for Arthur. His mean generalization score in this condition was 83%. Kim's data showed zero generalization in baseline also. Kim's data in the muiltiple exemplar condition indicated a five session period of no generalization, then she quickly reached criterion. Her mean score on this condition was 44%.

CHAPTER IV

DISCUSSION

The multiple baseline across tasks allowed adequate experimental control to affirm that: (a) there was an increase in acquisition of the probe items for Arthur as soon as training began in the multiple exemplar condition, (b) there was an increase of probe items for Kim as training began in the multiple exemplar condition, and (c) the training of multiple exemplar was functionally related to the increase of correct responses on the probe items given.

Therefore, it seems that the data for Arthur and Kim supported the hypothesis that performing a task with multiple exemplars would increase generalization across similar untrained tasks. Thus, it would seem that it would be good educational practice to train students with severe handicaps, using a variety of materials in order to achieve response in these students who tranditionally have difficulty in generalizing.

These results suggest important teaching and research implications, but must be cautiously interpreted in view of certain design limitations.

Teaching Implications

In order for the severely multiply handicapped student to develop into a more independent person, he or she must be able to perform responses across an array of natural situations where a response is appropriate and abstain from performance of that response where it is inappropriate. Horner & McDonald (1982) define this response as generalization. Horner & McDonald indicated that the easiest (most

common) strategy (to teach one good example of a skill) is ineffective in promoting generalization in students with severe handicaps. The student trained with a single example can be expected to learn a very restricted skill that will not facilitate competent performance across stimulus variations in the natural environments. The acquisition of multiple responses can occur with the selection of multiple training example that sample the range of stimulus variation to be encountered (Becher & Engleman, 1978; Horner & McDonald, 1982). Horner & McDonald suggest (for teachers who are already using multiple teaching exemplar or teaching in multiple settings) examining the range of stimulus variation the multiple exemplars cover. It is not enough to simply select many examples needed to teach an appropriate response to the various environmental stimuli.

The teaching of multiple exemplar response skills is important in vocational environments that require the student to make general capabilities to new tasks. The hinges, deadbolt locks, faucets, brackets, and handles are examples of applying multiple exemplar response skills on vocational tasks. The selection of the correct screwdriver, using wood, or assembly tasks require specific responses to different (often new) tasks for the student, therefore, requiring generalization skills. The field of prevocational education is founded on the assumption that skills acquired during prevocational training will be used effectively with new tasks in real vocational settings. Horner & McDonald go on to say that this assumption presumes that the learner will acquire multiple skills in the prevocational setting. Horner and McDonald found that it is unlikely if only one example of a

skill is trained that the learner will generalize to other response skills (regardless of how long the skill is used, once trained), and that multiple exemplar training is more effective than that of a single subject approach. Therefore, teachers in future planning should define settings (or stimulus conditions) in which the skill is expected to occur. Horner and McDonald suggest that the examination of these settings should result in specification of relevant stimulus variation and selection of training that sample this variation.

Another fact that a teacher should consider is the importance of controlling student errors. When designing instructional programs, teachers should select training exemplars both for relevance in teaching the desired skill, and for their appropriateness in avoiding common errors (Horner & McDonald, 1982).

Research Implications

The study provided an experimentally controlled analysis of generalized performance. As a result of the design, it can be said that the acquisition of generalized probe items was related to the multiple exemplar training. The strength of the design comes from defining the dependent variable as the level of prompting needed on the multiple exemplar items and the generalization probe items as the number of independent responses on a step of the task.

A critical assumption with this design is that performance on the three generalization probe items is representative of what the student's performance would be on the full display of examples (i.e., the several hundred other faucets, hinges, handles, deadbolt locks, or brackets not tested). Horner and McDonald (1982) point out an assumption like this

would require that the experimenter: a) provide an operational definition of the full display of stimulus situations in which the response (skill) is expected; and b) provide explicit description of the procedures used to select generalization probe items. The description would include a rationale for selection of stimulus characteristics considered relevant for correct performance along with a rationale for how selected probe items sample variations observed along these relevant characteristics (Horner & McDonald, 1982). The design will be more effective after it is applied with a variety of different dependent variables. The validity of the research design will increase when an additional set of generalization probe items are used under the same guidelines as the original set of generalization probe items were selected. This would then increase the assumption that the multiple exemplar training influenced the acquisition of the generalization probe items.

Limitations

The results in this study take a step toward defining teaching procedures and research implications needed for multiple exemplar response training toward severely handicapped individuals. The following two areas should be looked into before adequate interpretation can be made.

Trainer Bias. A single trainer conducted all of the multiple exemplar training and generalization probe sessions. The trainer was aware of the expected results of this study. It is possible that uncontrolled, inadvertant changes in trainer behavior may have affected student behavior. While the promptness and extent of improvement

following multiple exemplar training make this an unlikely possibility, the opportunity for trainer bias forces some reservation in interpreting the results (Johnson & Bolstad, 1973).

Single Environment The results from this study present a favorable argument for generalization across tasks (i.e., multiple exemplar to generalization probe items). These results were taken from a single vocational environment (setting) which leaves the assumption taken in this study to be somewhat insubstantial. The student needs to learn to generalize to other environments or settings since he/she will not always be in that same setting in the classroom, and especially not in a more natural vocational setting.

The two limitations mentioned above should be investigated to expand teaching procedures and research in the area of generlization skills. Even so, the data for this study justify a meaningful relationship between the use of multiple exemplar training procedures and acquisition that is functional across nontrained stimulus situations. The data in this study demonstrates experimental control of a generalized response.

CHAPTER V

SUMMARY

The study was conducted to investigate the generalization problem faced by severely multiply handicapped students. The severely multiply handicapped student has difficulty generalizing across trainers, materials, and settings. It was the intent of this study to determine the effect of supplying the student with multiple exemplars to promote generalization across untrained items.

Two students were included in the study. A multiple baseline was used to illustrate what effects training multiple items (metal faucet, metal hinge, and metal handle) had in promoting generalization across untrained items (large brackets, deadbold lock, and wood handles).

Both of the students, Arthur (Subject 1) and Kim (Subject 2), demonstrated an increase in the rate of acquisition of untrained items during the multiple exemplar condition. Arthur performed at a much faster rate of acquisition than did Kim, who had a more gradual rate of acquisition.

It can be concluded that the use of multiple exemplar training, with some students, promotes generalization across other training items. However, these results cannot be generalized to all individuals classified as severely multiply handicapped. This study is limited by population size, the number of exemplars used, and the single trainer bias.

REFERENCES

Becker, W. C., & Engleman, S. E. System for basic instruction: Theory and applications. In A. Catania & T. Brigham (Eds.), Handbook of Applied Behavior Analysis: Social and Instructional Process. NY: Irvington, 1978.

Brown, L., Nietupski, J., & Hamre-Nietupski, S. Criterion of ultimate functioning. In M. A. Thomas (Ed.). Hey, don't forget about me. Reston, VA: The Council for Exceptional Children, 1976, 2-15.

Brown, L., & York, R. Developing programs for severely handicapped students: Teacher training and classroom instruction. Focus on Exceptional Children. 1974, 6 (2), 24-36.

Drabman, R. S., Hammer, D., & Rosenbaum, M. S. Assessing generalization in behavior modification with children. The generalization map. Behavioral Assessment, 1979, 1, 203-219.

Holvoet, J., Mulligan, M., Schussler, N., Lacy, L., & Guess, D. The KICS model: Sequencing learning experineces for severely handicapped children and youth. Lawrence, KS: University of Kansas, Department of Special Education, 1982.

Horner, R. H., & McDonald, R. S. Comparison of single instance and general case instruction in teaching a generalized vocational skill. Journal of the Association for the Severely Handicapped, 1982, 7 (3), 7-20.

Johnson, S., & Bolstad, O. Methodological issues in naturalistic observations: Some problems and solutions for continued research. In L. Hammerlynck, L. Handy, & E. Mash (Eds.), Behavioral Change: Methodology, concepts and practice. Champaign, IL: Nescarch Press, 1973.

Kazdin, A. E. Methodological and assessment considerations in evaluating reinforcement programs in applied settings. Journal of Applied Behavior Analysis, 1973, 6, 517-531.

Lent, J. R., & McLean, B. M. The trainable retarded: The technology of teaching. In N. G. Haring & R. L. Schiefelbusch (Eds.), Teaching Special Children. NY: McGraw-Hill, 1976.

Sailor, W., & Mix, B. J. The TARC assessment system. Lawrence, KS: H & H Enterprises, 1975.

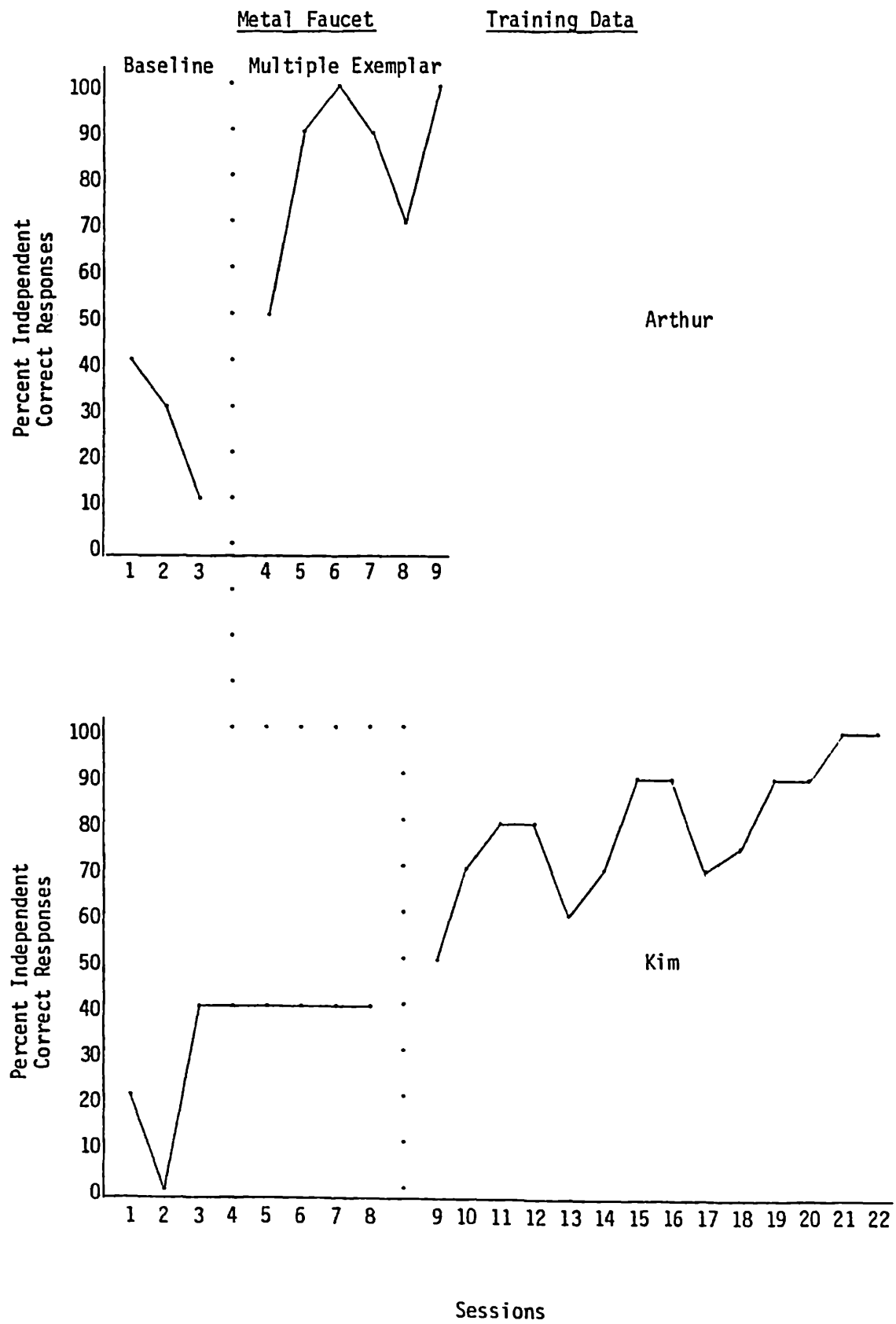
Schworm, R. & Abelseth, J. Teaching the individual with severe learning problems: Strategies which point to success. Education and Training of the Mentally Retarded, 1978, 13 146-153.

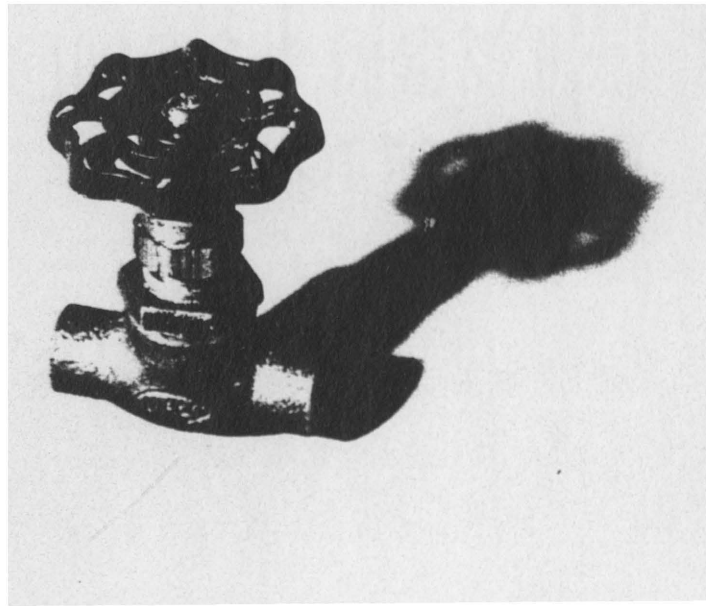
Stokes, T. F., & Baer, D. M. An implicit technology of generalization. Journal of Applied Behavior Analaysis, 1977, 10, 349-367.

Wehman, P., Abramson, M., & Norman, C. Transfer of training in behavior modification programs: An evaluative review. Journal of Special Education, 1977, 11, 212-231.

Appendix A: Data and Pictures of Training and Generalization Probe Items

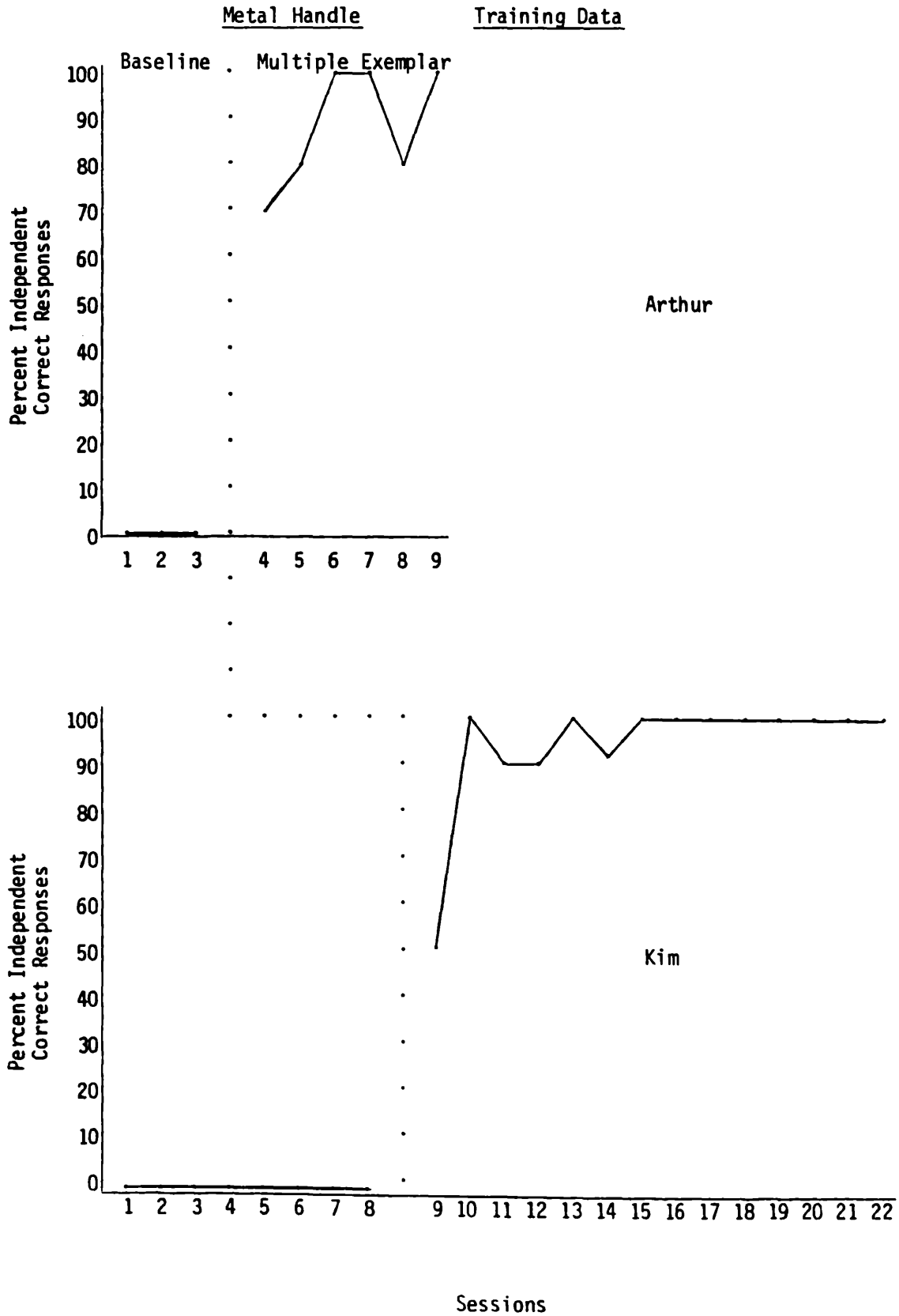
Figure 1. The percent of independent, correct responses made by the subjects on the metal faucet (a training task).

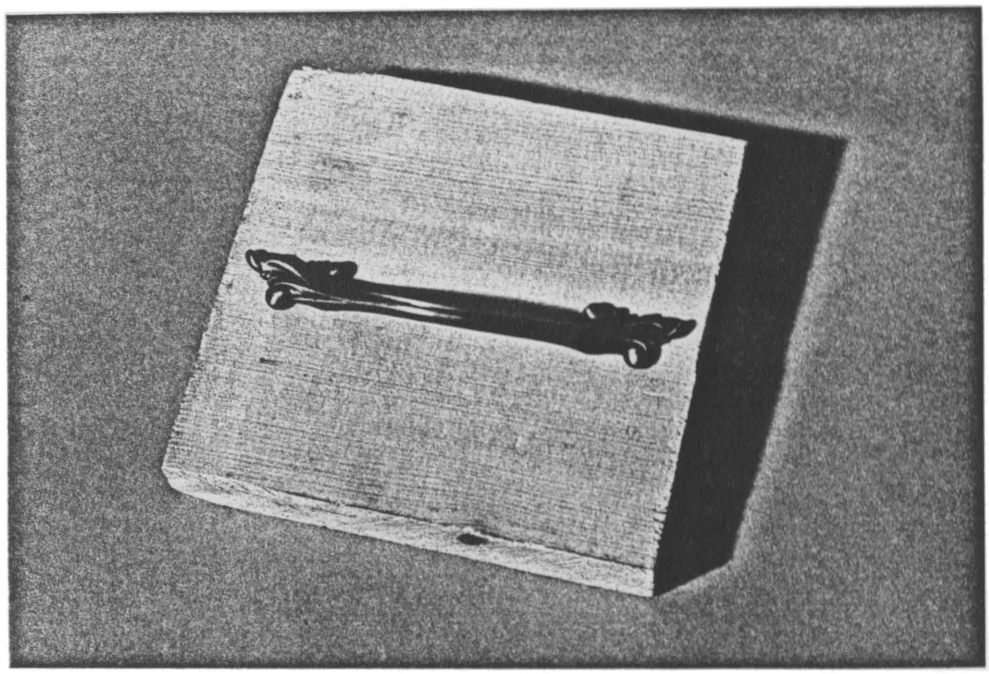

Metal Faucet
Training Data
Baseline
Multiple Exemplar
Percent Independent Correct Responses
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9
Arthur
Percent Independent Correct Responses
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Kim
Sessions



METAL FAUCET

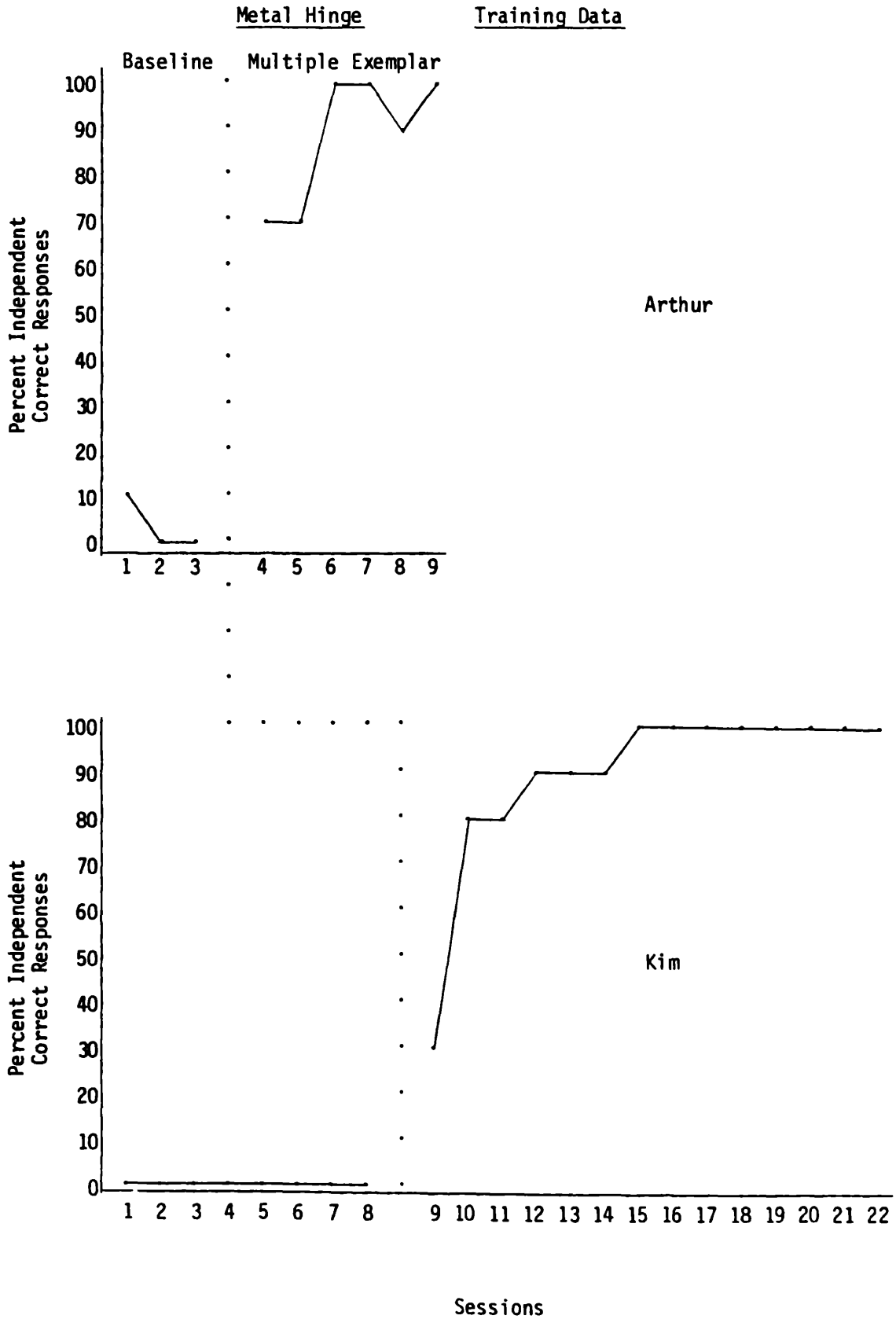
Figure 2. The percent of independent, correct responses made by the subjects on the metal handle (a training task).

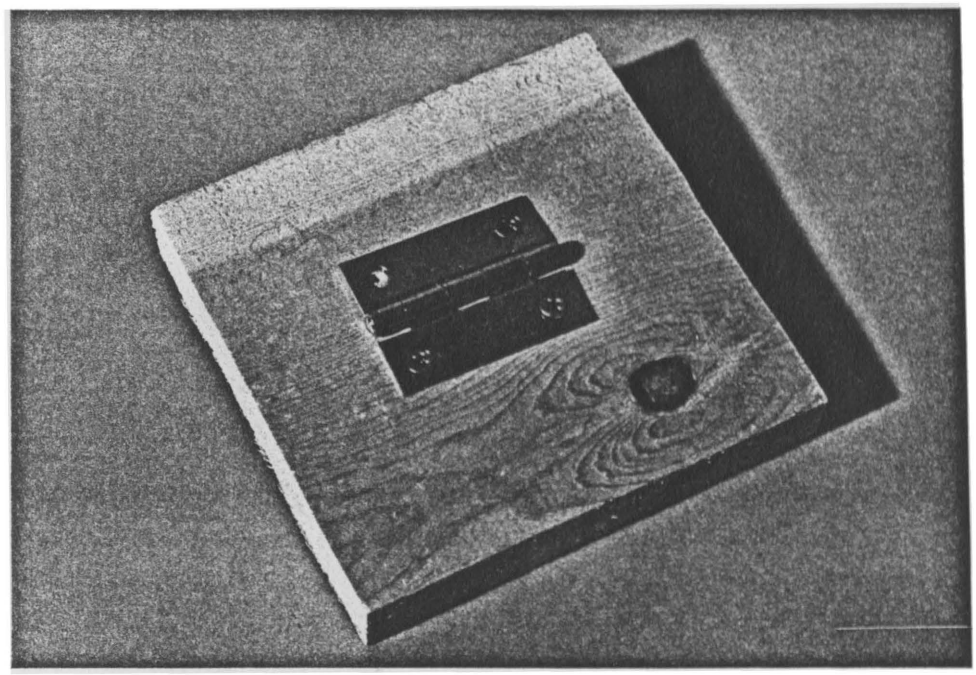

Metal Handle
Training Data
Baseline
Multiple Exemplar
Percent Independent Correct Responses
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9
Arthur
Percent Independent Correct Responses
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Kim
Sessions



METAL HANDLE

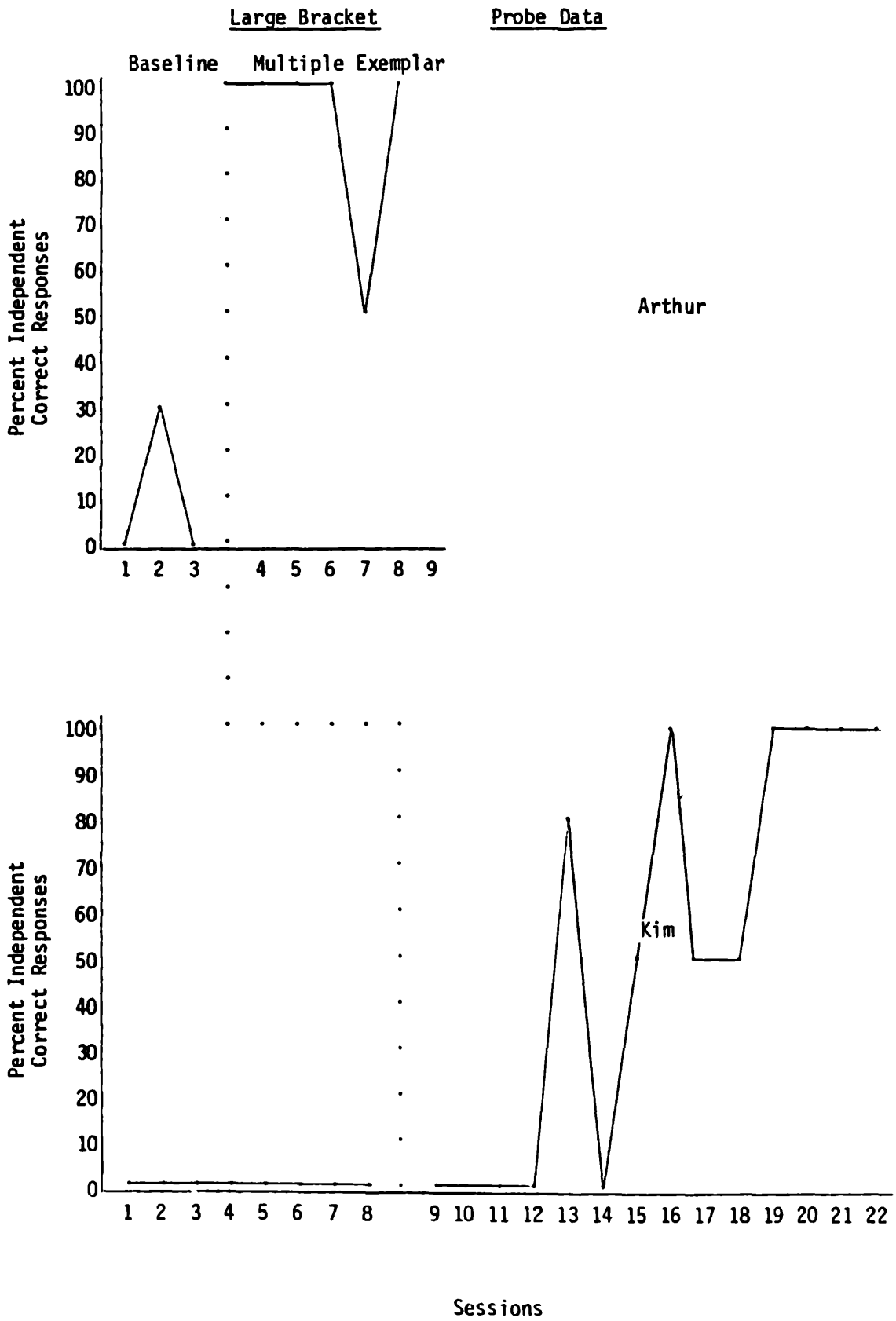
Figure 3. The percent of independent, correct responses made by the subjects on the metal hinge (a training task).

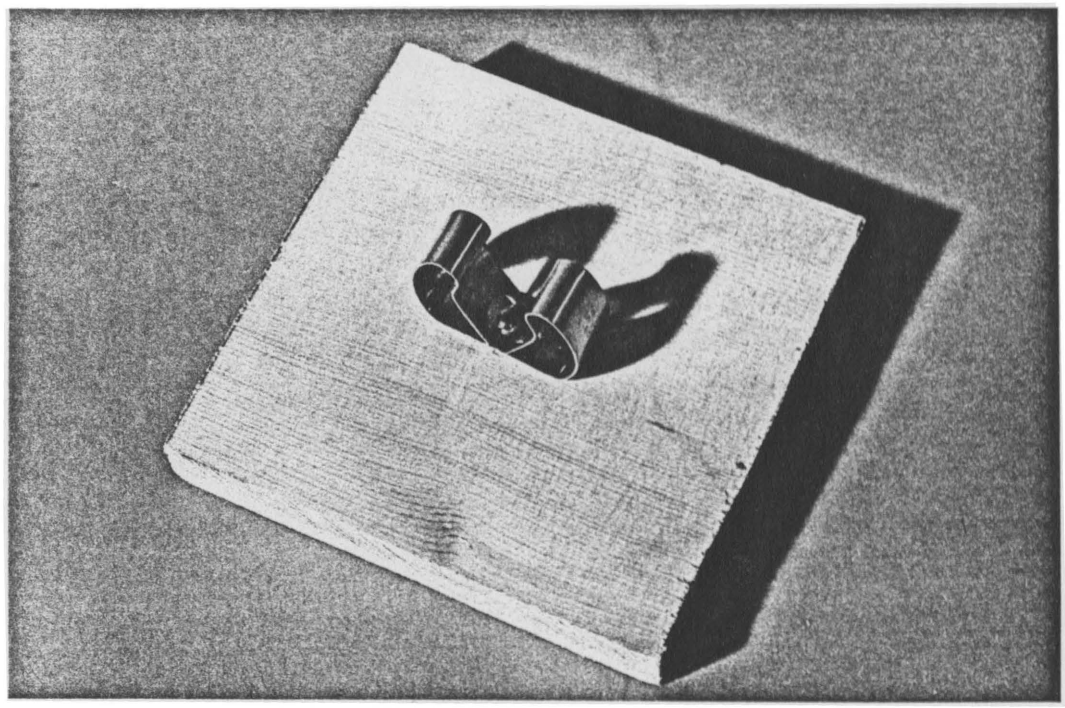

Metal Hinge
Training Data
Baseline
Multiple Exemplar
Percent Independent Correct Responses
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9
Arthur
Percent Independent Correct Responses
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Kim
Sessions



METAL HINGE

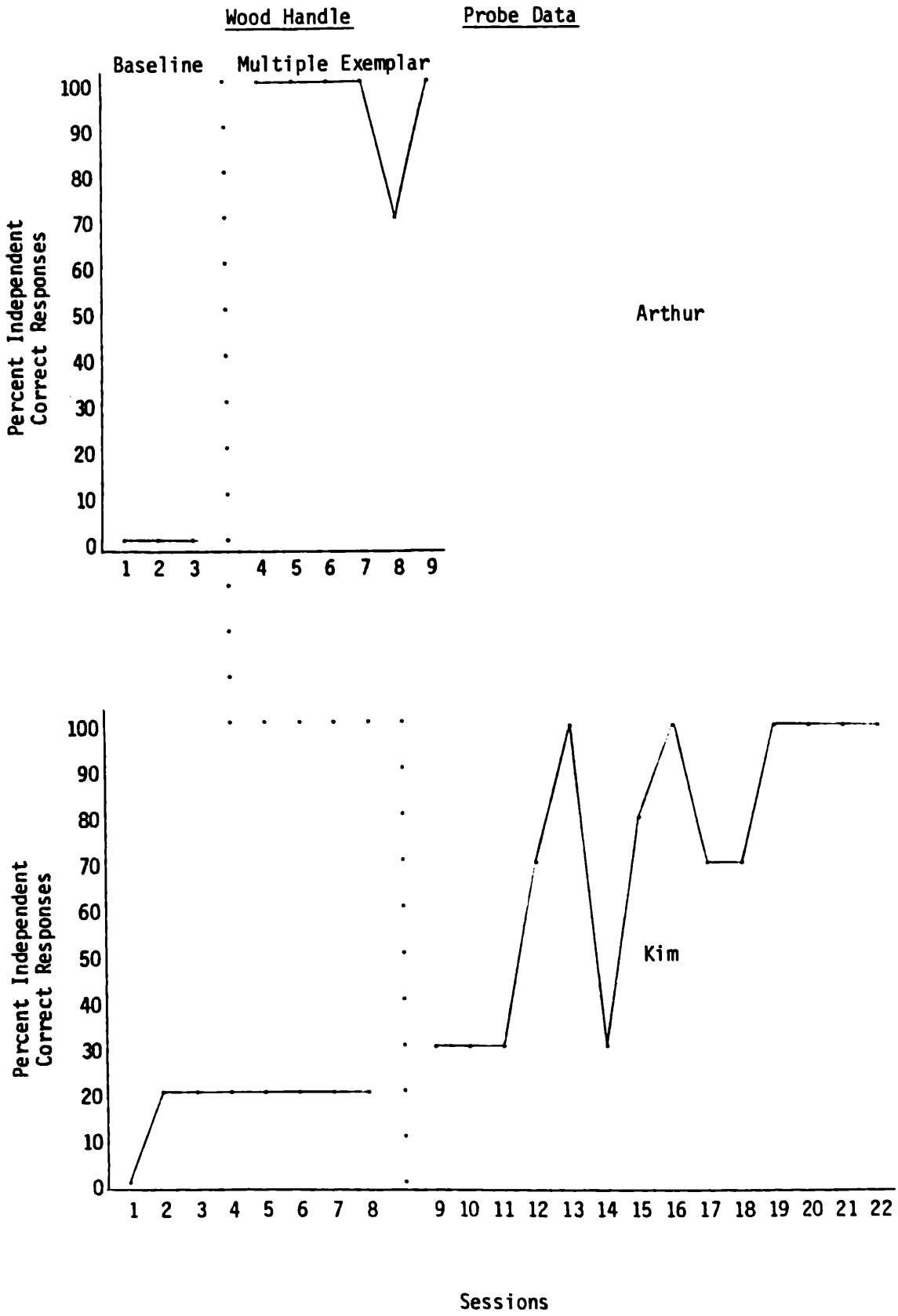
Figure 4. The percent of independent, correct responses made by the subjects on the large bracket (a probe item).

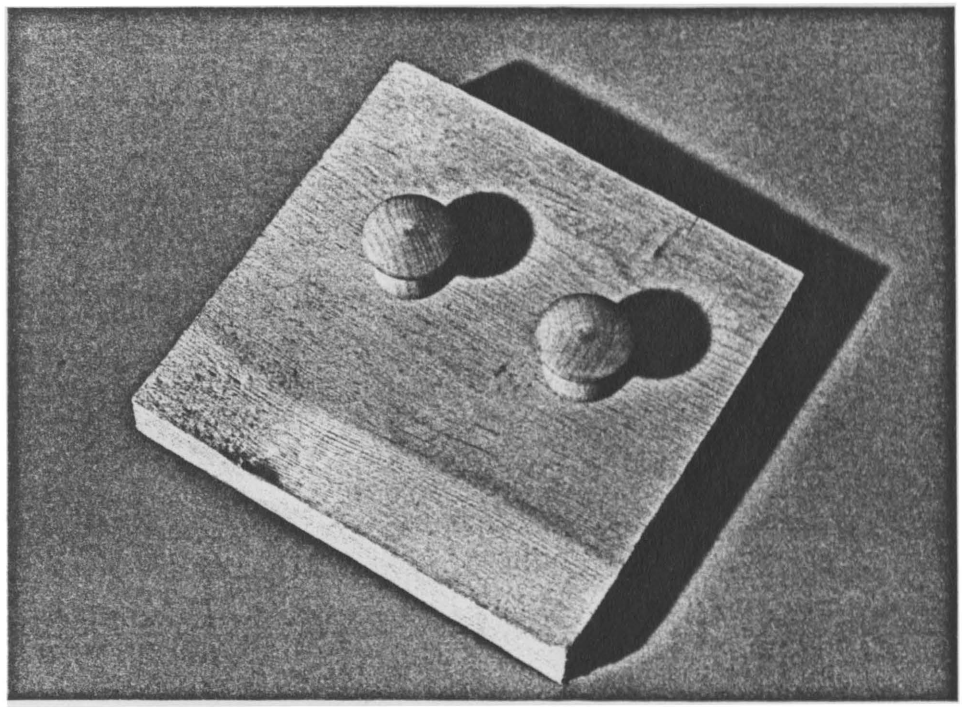

Large Bracket
Probe Data
Baseline
Multiple Exemplar
Percent Independent Correct Responses
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9
Arthur
Percent Independent Correct Responses
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Kim
Sessions



LARGE BRACKET

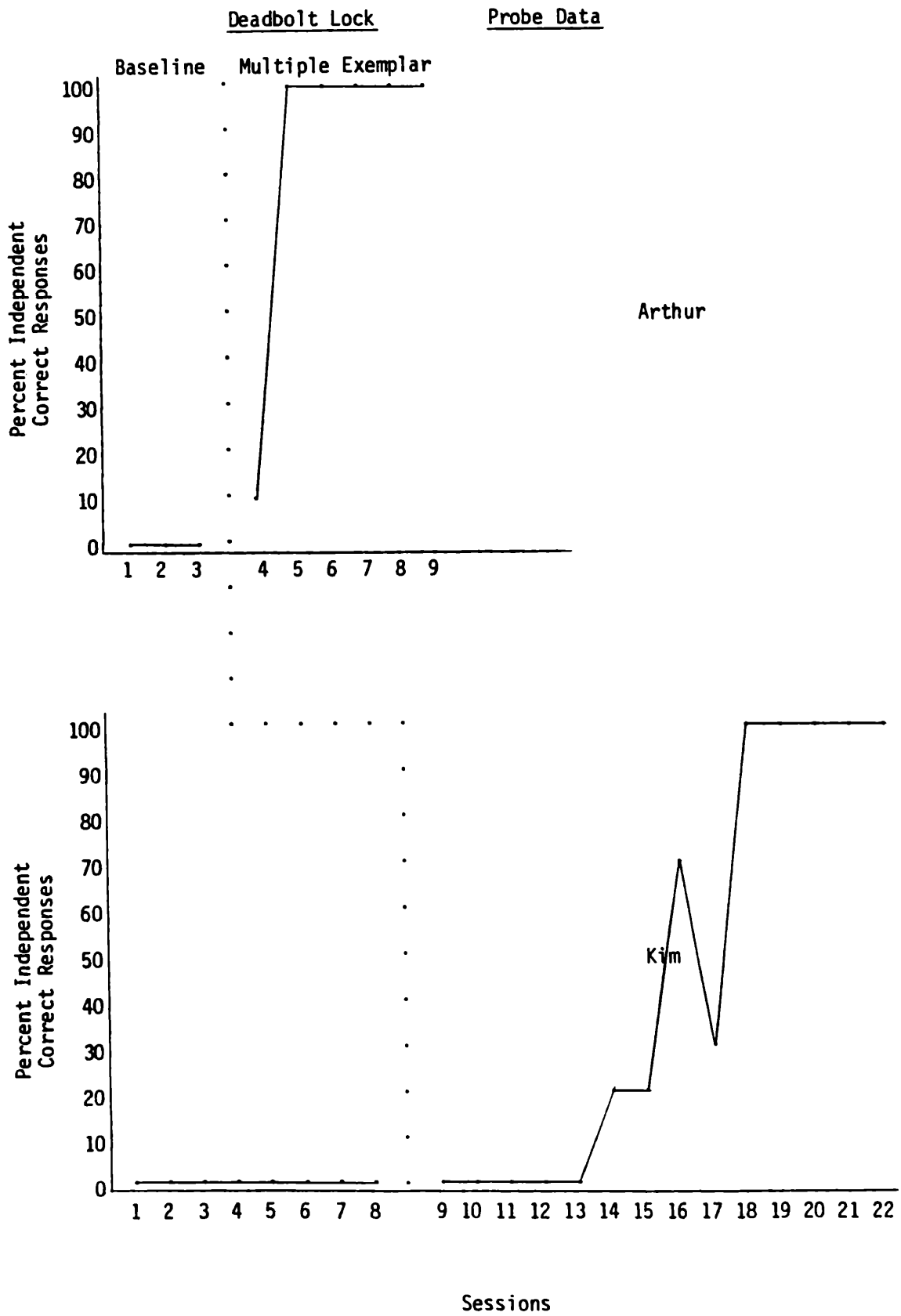
Figure 5. The percent of independent, correct responses made by the subjects on the wooden handles (a probe item).

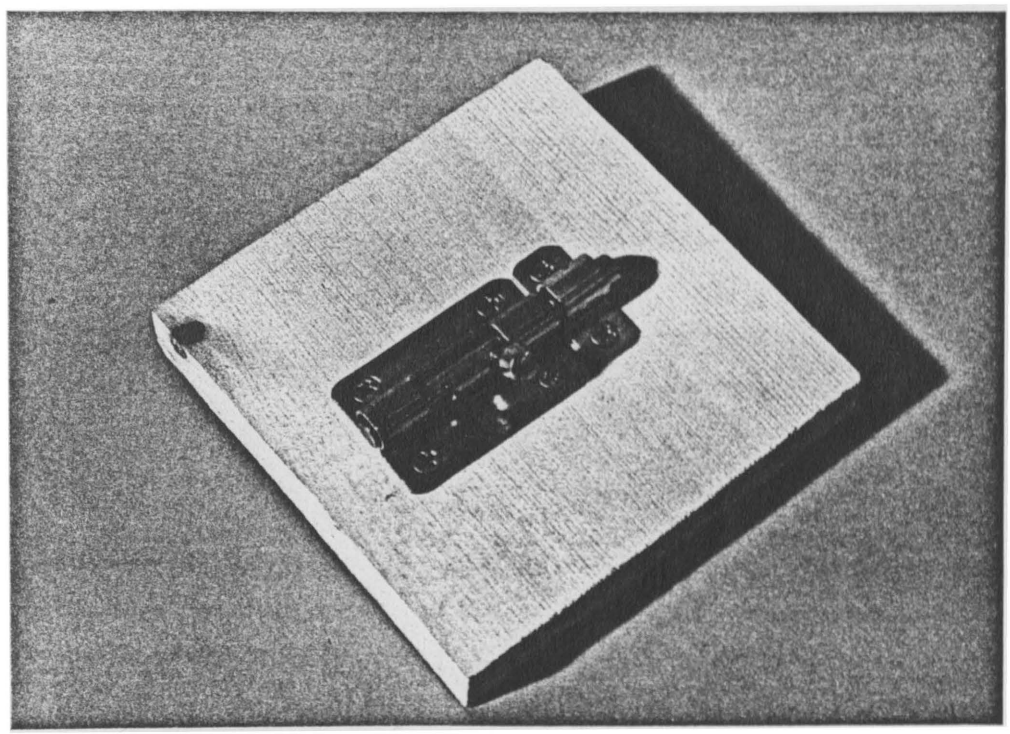

Wood Handle
Probe Data
Baseline
Multiple Exemplar
Percent Independent Correct Responses
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9
Arthur
Percent Independent Correct Responses
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Kim
Sessions



WOODEN HANDLES

Figure 6. The percent of independent, correct responses made by the subjects on the deadbolt lock (a probe item).


Deadbolt Lock
Probe Data
Baseline
Multiple Exemplar
Percent Independent Correct Responses
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9
Arthur
Percent Independent Correct Responses
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Kim
Sessions



DEADBOLT LOCK

Appendix B: Tables on Men Reliability Scores

Table 1

Mean Reliability Scores For Arthur

Item	Designation	Baseline	Multiple Exemplar
Faucet	Training item 1	90%	100%
Metal Handle	Training item 2	100%	100%
Hinge	Training item 3	100%	100%
Bracket	Probe item 1	100%	100%
Wooden Handle	Probe item 2	100%	100%
Deadbolt	Probe item 3	100%	100%

Table 2

Mean Reliability Scores For Kim

Item	Designation	Baseline	Multiple Exemplar
Faucet	Training item 1	100%	100%
Metal Handle	Training item 2	100%	100%
Hinge	Training item 3	100%	100%
Bracket	Probe item 1	100%	86%
Wooden Handle	Probe item 2	93%	100%
Deadbolt	Probe item 3	100%	100%