

*SOCIAL VALIDATION: THE EVOLUTION OF STANDARDS  
OF COMPETENCY FOR TARGET BEHAVIORS*

RONALD VAN HOUTEN

MOUNT SAINT VINCENT UNIVERSITY

The use of social validation procedures has become widespread in recent years. Although most researchers have used social validation procedures to select target behaviors and to evaluate whether the changes produced by a treatment program should be considered socially useful, little attention has been focused upon using the social validation process to determine the optimal levels for target behaviors. This paper suggests several ways in which social validation procedures can be employed in order to select when and how much to change target behaviors.

**DESCRIPTORS:** social validation, competency standards, normative data, goal selection, standard selection, response rate

Over the years, behavior modifiers have strived to increase the frequency of behaviors deemed desirable and to decrease the level of behaviors judged undesirable. Although considerable emphasis has been placed on evaluating the amount of change produced by experimental procedures, only recently have researchers directed their attention to assessing whether behaviors have improved sufficiently to consider a treatment socially useful (Kazdin, 1977; Wolf, 1978).

Basically, two methods have been employed to select target behaviors and to assess whether the changes produced by a treatment are of applied importance—social comparison and subjective evaluation. The first method involves comparing the performance of target subjects with that of peers who did not warrant treatment. If the client's behavior falls within the normative range of behavior, the treatment is considered successful (Patterson, 1974; Walker and Hops, 1976). The second method involves

having judges with special expertise rate the behaviors of subjects before and after the treatment is implemented (Briscoe, Hoffman, and Bailey, 1975; Kent and O'Leary, 1976; Maloney, Harper, Braukmann, Fixsen, Phillips, and Wolf, 1976; Porterfield, Herbert-Jackson, and Risley, 1976; Van Houten, Morrison, Jarvis, and McDonald, 1974).

The selection of appropriate target behaviors has been identified as an important problem in applied behavior analysis (Hawkins, 1975) and variations of the two above-mentioned procedures have proven increasingly useful in solving this problem. First, in the known groups method (McFall, 1976), the behavior of individuals comprising two groups that differ in some important global dimension are compared to identify which behaviors are uniquely correlated with the more successful group. Second, in the consumer satisfaction method (Wolf, 1978), clients or other consumers of our technology are asked to judge or rate various behaviors which are suspected to be important. Both methods rely upon experimental confirmation of the importance of the selected behaviors through the use of additional social validation procedures following treatment.

Although the selection of appropriate target behaviors is a valuable and necessary application

---

The author would like to thank Paul Nau and George Patterson for the many valuable suggestions they made regarding this manuscript. Preparation of this manuscript was partially supported by Canada Council Grant S74-1700-R2. Reprints may be obtained from Ron Van Houten, Psychology Department, Mount Saint Vincent University, Halifax, Nova Scotia, Canada, B3M 2J6.

of social validation techniques, it is only one step in the overall assessment process. Another step which has received considerably less attention is the selection of optimal levels of performance for the behaviors chosen in the previous step. This step becomes as important as the previous step if one considers that for most behaviors there exists a range of responding within which performance is most adaptive. When the limits of this range are unknown for a particular behavior, it is possible that one could terminate treatment when performance is above or below these limits. Hence, the behavior would not be occurring within its optimal range.

Terminating treatment below the lower limit of competent performance is clearly an undesirable treatment outcome. However, social validation carried out in the traditional manner would not provide information on whether or not this type of error had actually occurred. For example, if the terminal performance is superior to baseline performance a positive social validation is likely to result. Unfortunately the evaluation would not inform the researchers that higher levels of performance would lead to even more adaptive behavior on the part of the client. Conversely if the performance is not judged to be better following treatment, it is unclear whether the behavior selected is unimportant or if it was not changed sufficiently to yield positive results.

On the other hand, if treatment is terminated after the behavior is increased beyond the upper limit of competent performance, the resulting outcome could be equally detrimental to the client. An obvious example would be an under-assertive child who has been taught to make more requests. Initial increases in the frequency of requests might be considered increasingly adaptive; however, at a certain point, the behavior could begin to be considered pestering by those interacting with the child. At this point, the frequency of the behavior has surpassed its optimum value and would no longer be judged as desirable by those interacting with the child.

Social validation techniques employed in the traditional manner would not inform one whether the behavior was more functional at a lower level, nor could they inform one whether negative outcomes resulted from selecting the wrong behavior as a target or from producing too little or too much change in the level of the target behavior. Similarly, there is no guarantee that normative levels represent optimal performance.

In one experiment, Willner, Braukmann, Kirigin, Fixsen, Phillips, and Wolf (1977) found that youths judged that they preferred teaching parents who engaged in various behaviors such as joking, giving explanations, or providing offers to help. Although this study identifies various important behaviors, it does not tell practitioners what level or frequency of these behaviors is optimal. Just as it is quite possible that one could engage in too few of these behaviors, one could also engage in these behaviors too much. For example, it might be possible for an adult to joke or horse around too much, or to make too many offers to help, or to overdo giving explanations. Therefore, the next step in the evaluation of teaching parents may be to determine frequency aims for these behaviors. A similar argument could be made for identifying the situations where these behaviors are judged most appropriate. Still another step in the social validation process would be to identify the various interactions between these behaviors and their influence on treatment efficacy.

Failure to identify the limits of adaptive behavior makes it difficult if not impossible to determine how much to change a particular behavior in order to obtain the best results for the client. Another problem that may result from the failure to identify the optimal range of target behaviors is that practitioners cannot tell when to change a particular behavior and when to leave it alone. Without the above-mentioned information, it would be difficult to determine whether a treatment which produces a socially validated improvement in one individual's per-

formance would also produce such a change in another individual's performance since one would now know if that behavior is already within the optimal range.

In order to know when to initiate and terminate a treatment, practitioners require socially validated standards for which they can aim. The availability of these standards is especially important to those who are not themselves engaged in applied research, since these individuals are often not in a position to perform a complex social validation of the changes they produce because of financial or temporal constraints.

This article suggests several ways in which social validation procedures can be employed to determine these aims: normative based selection procedures and experimentally based selection procedures.

#### *Selecting Standards of Competence through Normative Based Selection Procedures*

One way of determining socially validated goals is to assess the performance of individuals who are judged to be highly competent in the area of interest. The norms obtained from these individuals can then serve as performance goals for less competent individuals. Setting norms in this manner utilizes both commonly employed methods of social validation since it involves first selecting individuals who are universally judged to be competent at a given behavior and second determining the level of their performance.

The first decision to be made involves choosing whether to obtain a single universal goal for each task from the most competent individuals we can find or to obtain separate goals from each special population such as the developmentally retarded or the handicapped. In the long run, it is probably better to select a single goal from those judged most competent within the society. Although goals obtained by this method may at first appear overly stringent for some populations, they can still serve as a yardstick against which changes produced by any

treatment can be assessed. Furthermore, they possess the added advantage of serving as a challenge to develop more powerful change techniques. A failure to bring an individual's behavior within the competent range does not necessarily imply an experimental failure since it is certainly important to bring individuals as close as possible to an optimal level of performance. However, researchers will remain aware of the possibility of producing even more adaptive levels of performance. Although it is possible to avoid this issue entirely by obtaining separate competency based norms for every population, this is probably unwise since the levels of behavior which indicate competence do not change because some individuals have more difficulty obtaining them. However, it may be practical to set lower short-term goals in some instances because techniques may not presently exist to bring some individuals' performance within the component range.

An advantage of using standards based on the performance of competent individuals is that doing so does not limit potential for improvement. One objection that has been raised about using norms obtained from untreated peers as a treatment standard is that these norms themselves could be improved in many instances (Kazdin, 1977). For example, in the classroom situation it may be possible to accelerate the normative level of academic performance in most classrooms. This objection does not apply to normative data obtained from competent individuals because these data are based on the performance of the most competent individuals; and, therefore, it may be quite appropriate to accelerate the performance of an entire group to these levels. Lastly, though norms based on the performance of competent individuals lack some of the disadvantages of group norms, they should not be considered completely inflexible because additional studies may indicate higher levels of competence are possible. In fact the selection of target behavior and performance standards are both best viewed as an iterative

process where repeated social validation applied to a particular problem leads to the improved delineation of important treatment goals.

### *Obtaining Standards from Competent Individuals*

It is quite simple to obtain rate and accuracy standards for typists or stenographers. Secretarial schools have been employing such standards for many years. Merely being capable of performing a task such as typing does not demonstrate that an individual is a proficient typist. For example, one typist might be able to type 100 to 120 words per min without errors while another can only type between 30 and 40 words per min without errors. Since both students can type without errors, should we consider that they are equally proficient? In order to answer this question we must know what the term proficiency really means. Typists are proficient only when they can type at a high enough rate for their services to be useful to others or to themselves. Most individuals could type a perfect page given enough time, but that does not make them proficient typists.

If we select standards for typing proficiency by examining the performance of randomly sampled adults, it is unlikely that the standards obtained would represent maximally useful performance. Even if we randomly sampled the performance of secretaries, the resulting standards would only represent mediocre performance. Only by sampling the performance of secretaries rated as competent typists by their employers could we be sure of obtaining a standard which reflects competence, and which a powerful behavioral technology should aim to produce. Hence, one way of selecting competent individuals is to ask those who require the behavior at what level they judge it to be functional.

Similarly, standards in mathematics and reading could be easily obtained by looking at individuals who are judged to be highly competent at various mathematics and reading skills, while standards for essay writing could be ob-

tained by looking at the performance of the best essayists. Therefore if we wished to obtain standards for adults who are learning to write essays we would get them from established writers who are judged competent by our society. These standards would then serve as goals for both adults and children who are learning to write. Again although these standards may seem overly stringent when applied to children, we should remember that a competent typist is competent regardless of age as is a competent musician, reader, or writer. The fact that there may be fewer children than adults who are competent at a particular task does not imply that the ultimate goal for children's performance is any different from that of an adult's.

For example, if we found that a competent adult could solve approximately 80 single digit multiplication problems per minute without any errors, we could set this as a goal for children and adults alike. Similarly, if we found that competent adults could read material at their reading level aloud at a rate of 150 words per min with 98% accuracy, then we could presume that children learning to read should be as competent when they meet this standard on their reading material. Although the level of the material being read may be different, the reading rate and accuracy which indicate competence may remain the same. Furthermore, it is undoubtedly true that children often become as competent as adults at specific tasks.

In both of the above examples, some individuals might have difficulty obtaining these standards, but movement in the direction of the standard would often be judged successful. Applying rigorous standards may mean we would have at least some idea of the degree to which we could approach the level of performance of a competent model, thereby revealing how much our technology needed to be improved.

If we instead derive performance standards from normative data we run the risk that the reading and mathematics performance of the "average" student is below the competent range. If we adopt these standards, we will likely termi-

nate treatment before students are competent at a task.

The inadequacies of normative standards obtained from the overall population are excellently illustrated by the data obtained by White (1975) and Thomas, Presland, Grant, and Glynn (1978) on natural rates of teacher approval and disapproval. White measured the natural levels of approval and disapproval in sixteen classrooms. An analysis of her results indicates that the mean rate of teacher praise was .36 per min across all grades. Furthermore, teacher praise rate declined with increasing grade level so that high school teachers only praised someone every five or ten minutes. Thomas *et al.* (1978) examined the teacher praise rates of seventh grade teachers and found that the average teacher praise rate was .20 praises per min. From what is currently known about optimal rates of teacher praise (Cossairt, Hall, and Hopkins, 1973; Hall, Lund, and Jackson, 1968; McAllister, Stachowiak, Baer, and Conderman, 1969), it is unlikely that these rates would represent good standards for applied behavior analysts working with problems in the public schools. Alternatively, Nelson and Bowles (1975) suggest that normative data could be obtained from teachers described as highly competent. Although there is no guarantee that these goals will prove adequate, they can serve as a best guess in the first step of the social validation process.

By using competent behavior as a standard, we de-emphasize some of the relativity involved in norm selection because we provide goals which are judged to be functional. Furthermore, our goals are based upon estimates obtained from individuals who are performing the task at a useful level.

*Selecting competent models.* One problem associated with determining normative based competency standards is deciding which individuals are competent at a given task. As a first step one can ask those most often involved with rating the adequacy of the target individual's performance. We could ask teachers, parents,

employers, or others who are in a position to rate an individual's performance. For example, if we wanted to establish standards for reading, we could ask the superintendent of schools in which schools students read the best. Next the teachers in these schools could be asked to identify their best readers. The resulting standards could be considered a best guess estimate of competent performance. Further research and social validations could then help to establish how good these standards really are. Again successive application of the social validation strategy should lead to more refined standards.

*Selecting the correct measure.* Another problem associated with determining standards of competent behavior from normative data is that we must select the most appropriate measure upon which to base our assessment. If we select an incomplete or incorrect measure, the standards obtained may not bear any relationship to the quality or social utility of the performance. For example, if in our mathematics example we had selected percentage correct alone as our standard, we would not be able to distinguish between a student who could only work 20 problems per min with no errors (100% correct) from one who could work 80 problems per min with no errors (100% correct). Clearly the addition of rate standards allows us to differentiate between a competent individual capable of working 80 problems from a less capable person who could only work 20.

Similar problems could easily arise when examining a complex task like storywriting. For example, a teacher may observe that the percentage of different adjectives used by the students is far lower than that used by competent writers. Therefore the teacher might establish a program to increase the percentage of different adjectives used. However, even if the program is successful, the teacher cannot be sure that students who increase their use of adjectives are using them in the same way as good writers. Students might string together many adjectives of the same kind, such as "the red, yellow, and blue house" rather than distributing adjectives

or using different kinds of adjectives, such as the "old, deserted, grey house." It is easy to see how a general measure like the percentage of different adjectives could also be an oversimplified measure.

This does not imply that it is not possible to set objective competency standards but it does mean that we must specify our variables so that they will be related to social validation measures, which in the case of storywriting might be overall quality. If we can identify all the relevant features of good adjective usage then it will be possible to place consequences on each aspect and bring them into the normative range of competent individuals. In the case of adjective use, some important variables might be diversity in the use of different types of adjectives and the way adjectives are distributed among the various nouns. However, only when the most relevant attributes of good adjective use are clearly identified can we begin to teach students to improve systematically. A similar argument could be made about setting standards for optimal levels of teacher praise. It would be easy to conclude that frequency of praise is important when in fact a closer analysis might reveal that whether praise is descriptive or accompanied with a smile or a pat also influences whether it is effective.

Although the relevant dimensions of many behaviors may be quite complex, it is still possible to move toward the establishment of performance standards because of the iterative nature of the social validation process. Each successive study can yield a closer approximation of relevant aspects and standards.

#### *Selecting Standards of Competence through Experimentally Based Procedures*

Another way of assessing and refining socially validated goals is through two methods of empirical assessment. Using the first method, the experimenter manipulates the behavior of interest over its entire range and determines at which values the behavior is maximally useful or effective in attaining important functional goals. The aspect of this procedure that sets it

apart from the more traditional application of social validation procedures is that the behavior of interest is varied over a wide range of values in order to select a range of values within which performance is optimal. Clearly this approach is quite compatible with that employed in normative based selection procedures. Social validations which experimentally vary the frequency of behaviors believed to be important because they are consistently found in the repertoire of individuals judged to be competent at a particular task can lead to the empirical validation of these behaviors. Unfortunately few researchers have performed an experimental validation of competency standards; therefore, few examples exist typifying this approach.

In one example, Warren, Rogers-Warren, and Baer (1976) taught preschool children to increase their offers to share materials with their peers. In the case of sharing, the authors concluded that offers to share were most appropriate when they were accepted, and that if a large number of offers to share were rejected, then these offers were functionally inappropriate to the children receiving them. One factor which the authors suspected might influence the likelihood that an offer would be accepted was the rate of offers to share. In two experiments the authors increased the frequency of offers to share while examining the percentage of offers accepted. They found that the percentage of offers accepted declined as the frequency of offers to share increased. These results implied that an inverted U-shaped function exists such that as offer rates increase, the percentage of offers accepted increases up to a certain point and then declines. The authors concluded that if the frequency of certain behaviors determines whether individuals will react in a positive or negative way to them, then the target rate or goal should be chosen according to these results. The results of this study illustrates how precisely treatment goals must be specified. If practitioners interested in increasing the sharing behavior of a group of children were not aware of these findings, they might teach some children who ini-

tially have effective offer rates to respond at a level which would be socially less useful than their baseline rate.

In another example, Jones and Azrin (1969) found that they could virtually eliminate stuttering by having clients synchronize their speech to a regular vibrotactile stimulus. Although this procedure markedly suppressed stuttering at beat durations less than 5 sec, the resulting speech did not always sound natural. In order to determine how beat duration influenced "naturalness" of speech, Jones and Azrin taped 1-min segments of the clients' speech at nine different beat durations. These samples were then played to 12 judges who made independent judgments about the naturalness of each 1-min segment of taped speech. The results indicated that a beat duration of 1 to 3 sec produced the most natural sounding speech in all four clients. Data also indicated that the number of words read per minute were also at or near their peak value within these durations. It is interesting to note that the performances of all four clients were essentially similar. On the basis of the data obtained on stuttering, naturalness, and reading rate, it was possible for Jones and Azrin to conclude that an effective way to use a "metronome procedure as a clinical technique would be to have a stimulus duration of about 1 to 3 sec and a stimulus-off duration of 1-sec." It is interesting to note that in this example the experimental analysis of the effects of beat duration on stuttering produced compatible results on all validation measures. If all three measures had not been in such close accord, it would have been necessary to set priorities and trade each measure off against the others.

Similarly, communication skills, public speaking techniques, and other behaviors could be evaluated by their effects on their audience. For example, if researchers wanted to examine the effects of eye contact or gestures on public speakers, they could vary eye contact while keeping all other variables constant. If a particular range of durations were found to be optimally effective, clients could be taught to perform within

this range. Comparable results could be easily obtained for loudness of voice and other factors which might influence the efficiency of public speakers. Subsequent analysis performed on other variables would help identify many important effects and interactions. No single researcher would be expected to vary all suspected conditions; however, over a period of time more knowledge should be accumulated through a series of systematic replications. The end results of this operation would be empirically determined performance goals. The major difficulty with determining goals in this manner is the expense in time and effort associated with varying behaviors over their entire range of occurrence. However, in the long run this problem is offset by the degree of confidence one can assign to standards obtained in this way.

The above examples illustrate how the optimal range of performance can be selected for a particular behavior by varying its frequency of occurrence over a wide range of values. A second way of evaluating the adequacy of performance standards is to examine the effects of training to various standards on the subsequent acquisition of more complex skills. Clearly this approach is relevant whenever the behavior of interest is also a prerequisite for other important behaviors which must also be acquired.

For example, Van Houten (*in press-a*) examined the relationship between the rate at which students could complete basic single digit multiplication facts (of the form  $7 \times 8$ ) and their performance on long multiplication and division problems. The basic measure in this experiment was the number of long multiplication and division problems completed correctly each day on a 10-min test. During some conditions students were drilled on their basic multiplication facts at another time of the day, and during other conditions they were not. The results of this study revealed that students' performance on the long multiplication and division tests only improved during conditions when students were given drill on basic multiplication facts. Whenever multiplication drill was re-

moved, performance on the test problems neither improved nor deteriorated but rather remained the same. It should be noted that students rarely made errors during multiplication drill from the onset of the study. Therefore, a teacher who only considered accuracy on basic multiplication tables to be important would have concluded that these students had learned their multiplication tables and therefore required no further training. However, a teacher who also examined their work rate would have noted that they could only complete between 30 and 40 problems per min at the beginning of the experiment and between 60 and 70 problems per min at the end of the experiment. This difference becomes behaviorally significant when one considers that performance on this task was positively correlated with improvement on the more complex test problems. It is possible that we could terminate training on basic number facts such as the addition or multiplication table or printing before students have obtained a performance level sufficient to allow them to improve on more complex tasks for which these skills are a prerequisite.

Haughton (1972) has suggested that a similar relationship exists between the rate at which a child can print or cursively write, and spelling and composition writing. A similar approach can be employed in the evaluation of social behaviors. For example, Ayllon and Azrin (1968) suggested that practitioners teach behaviors that would be functional for individuals in their environment when they were reintegrated into the community. In one study, Fixsen, Phillips, Phillips, and Wolf (1976) evaluated the long-term validity of the target behaviors they selected with predelinquent youths by examining subsequent effects on school performance and recidivism. Their data validated their choice of target behaviors because they produced a marked reduction in the number of police and court contacts per youth up to two years following treatment. Positive results were also obtained with various aspects of school performance.

### AN EXAMPLE OF GOAL SELECTION WITH STORYWRITING

The first portion of this paper has been concerned with the general issue of selecting treatment goals or standards. The following discussion illustrates how the problems of goal selection can influence a program designed to teach children to write better stories. It also illustrates how the social validation process functions to develop standards through successive approximations.

In recent years, several articles have been published that suggest that judges give improved ratings of story quality when elementary students are taught to increase the number of different action words used in their stories (Ballard and Glynn, 1975; Maloney and Hopkins, 1973). However, Van Houten (*in press-b*) was unable to replicate these findings with elementary school children in a carefully controlled study where the number of different action words used during a 7-min story was increased from a baseline level of 4.7 to a posttreatment level of 12.1 words. When judges rated these stories on the dimension of overall quality, no differences were detected between stories written during the baseline and treatment conditions.

Similarly, Van Houten and McLellan (unpublished data) increased the number of different action words written by secondary school students during a 5-min writing period from a baseline rate of 8.9 to a posttreatment level of 19.7 words. Again, quality ratings made by experimentally naive judges revealed no changes in story quality. It is clear that these results are not in accord with the often replicated finding that increases in the number of action words written in a story are associated with increased story quality ratings.

At first it may seem confusing that changes in different action word usage were not associated with changes in judged story quality. However, if one compares these studies using an equivalent measure, a certain degree of orderli-



ness begins to emerge. One alternative is to examine the percentage of different action words used in each story. This measure would be superior to that of rate of different action word use because it is not confounded by the number of words written. Hence students could not increase the percentage of action words used by writing more rather than by changing their writing style.

However, even when the Van Houten (*in press-b*) and Van Houten and McLellan data were corrected for overall rate, it was still clear that the treatments had an effect on action word use. For example, the percentages of different action words used during the pretreatment condition were 13% and 11.2% for the elementary and secondary students, respectively. Posttreatment percentages were 15.2% and 18.3%, respectively. Clearly the treatments produced an increase in the percentage of action words in both studies. Since changes in the percentage of different action words were not associated with increases in story quality, it would be interesting to compare these levels with normative based competence standards obtained from writers who are universally judged as competent.

Therefore random segments from established authors which were of similar length to the stories written by these students were examined. Ten excerpts were randomly selected from the short stories of renowned writers (Cline, 1952). These were Anton Chekhov, Steven Crane, Joseph Conrad, D. H. Lawrence, James Joyce, Somerset Maugham, and John Steinbeck. The length of these excerpts, 80 words, was selected in order to keep length roughly comparable to that of the elementary and secondary students' stories.

The mean percentage of different verbs used by these authors was 11.3% with a range of 10% to 13%. Clearly, the percentages calculated from the students' stories during the baseline condition fell within this range. Similarly, 80-word segments taken from award winning stories of elementary and secondary students contained 12.2% different action words. This suggests that increasing student action word rate

did not produce a change in story quality because performance on this measure was already within an optimal range when the experiment began. It is also possible that in previous studies where experiments reported having improved stories by increasing the number of different action words used, students were highly repetitious in their use of action words during the baseline condition and hence were well below the percentages obtained from competent writers. If this were the case, any treatment which caused them to use action words in a way more similar to that of good writers would be associated with a judgment of improved story quality.

Unfortunately, it is impossible to compare the percentages in many of these studies because the authors did not report percentage data (Ballard and Glynn, 1975; Maloney and Hopkins, 1973) nor did they present sufficient information to allow for its computation. For example, Maloney and Hopkins (1973) gave students up to 40 min to write a 10-sentence story. Tokens were then awarded for increases in the number of different action words used per sentence. However, no information was provided on the number of words written or on sentence length, and hence it was not possible to calculate the percentage of action words per story. However, a rough comparison is possible. They reported that during the baseline condition the children used .39 different action words per sentence and during the treatment condition they used .98 different action words per sentence. If we assume that these children wrote sentences that averaged seven words or more and that this number was constant throughout the study, then the percentage of action words changed from a maximum of 6% to a maximum of 14%. If these estimates are correct, the baseline performance in this instance would be far below the norms obtained from competent writers of all ages.

In another study, it was possible to calculate the percentage of different action words directly from the experimental data (Maloney, Jacobson, and Hopkins, 1975). In this study, a teacher

had third-grade students write stories that were five sentences long. During a baseline condition, action words accounted for 4% of the total; during the treatment condition (when different action word usage was reinforced) the percentage increased to 13%. Further, judges rated the stories written during the condition when different action word usage was reinforced as superior to those written during the other conditions. In this study, the authors increased the percentage of different action words from a level far below that used by competent writers to a level within the range used by competent writers and obtained improved quality ratings.

These studies illustrate how performance standards can be evolved that can help predict whether or not a student will benefit from a treatment designed to increase the level of a behavior. Although increasing the percentage of different action words in a story can clearly lead to increases in a socially validated measure (story quality) if students are below competent norms, it is not likely to benefit students who are already responding within these norms. Hence knowledge of these norms enables practitioners to determine more carefully when a treatment will have a socially validated effect.

#### *Social Validation as an Iterative Process*

The above-mentioned example illustrates why social validation is best viewed as an iterative process which when properly employed helps researchers converge on the best treatment targets along with appropriate treatment aims or levels for selected behavior. Finally, this process narrows down the specific situations in which these behaviors are most appropriate. The outcome of the social validation process is the selection of appropriate target behaviors along with competency aims expressed in terms of optimal frequency, duration, accuracy percentage, or latency, and the delineation of the most appropriate settings for these target behaviors to occur. The process is comprised of normative assessments, subjective judgments, and global short- and long-term validations. The process

often begins by selecting behaviors and estimating competent levels of these behaviors through gathering normative data on the most competent individuals or by comparing the normative behavior of known groups of competent and incompetent individuals (McFall, 1976). Next researchers obtain closer and closer approximations of the best target behaviors and standards through the two types of experimental social validation procedures. First, behaviors believed to be important are varied across their natural range of occurrence and the effect on the environment and the rating behavior of judges are noted. Second, the global or long-term effects of changing specific target behaviors are noted.

The experimental aspects of the social validation process enable researchers to converge on closer approximations of guidelines specifying which behaviors are good candidates for change as well as when and how much these behaviors should be changed.

#### REFERENCES

- Ayllon, T. and Azrin, N. *The token economy*. New York: Appleton-Century-Crofts, 1968.
- Ballard, K. D. and Glynn, T. Behavioral self-management in story writing with elementary school children. *Journal of Applied Behavior Analysis*, 1975, 8, 387-398.
- Briscoe, R. V., Hoffman, D. B., and Bailey, J. S. Behavioral community psychology: training a community board to problem solve. *Journal of Applied Behavior Analysis*, 1975, 8, 157-168.
- Cline, C. L. *The Rinehart book of short stories*. New York: Holt, Rinehart & Winston, 1952.
- Cossairt, A., Hall, R. V., and Hopkins, B. L. The effects of experimenter's instructions, feedback, and praise on teacher praise and student attending behavior. *Journal of Applied Behavior Analysis*, 1973, 6, 89-100.
- Fixsen, D. L., Phillips, E. L., Phillips, E. A., and Wolf, M. M. The teaching family model of group home treatment. In W. E. Craighead, A. E. Kazdin, and M. J. Mahoney (Eds), *Behavior modification: Principles, issues, and applications*. Boston: Houghton Mifflin, 1976.
- Hall, R. V., Lund, D., and Jackson, D. Effects of teacher behavior on study behavior. *Journal of Applied Behavior Analysis*, 1968, 1, 1-12.
- Hawkins, R. P. Who decided that was the problem? Two stages of responsibility for applied behavior

- analysts. In J. S. Wood (Ed), *Issues in evaluating behavior modification*. Champaign, Illinois: Research Press, 1975.
- Haughton, E. Aims growing and sharing. In J. Jordan and L. Robbins (Eds), *Let's try doing something else kind of thing: behavioral principals and the exceptional child*. Reston, Virginia: The Council for Exceptional Children, 1972.
- Jones, R. J. and Azrin, N. H. Behavioral engineering: stuttering as a function of stimulus duration during speech synchronization. *Journal of Applied Behavior Analysis*, 1969, **2**, 223-229.
- Kazdin, A. E. Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification*, 1977, **1**, 427-452.
- Kent, R. N. and O'Leary, K. D. A controlled evaluation of behavior modification with conduct problem children. *Journal of Consulting and Clinical Psychology*, 1976, **44**, 586-596.
- Maloney, D. M., Harper, T. M., Braukmann, C. J., Fixsen, D. L., Phillips, E. L., and Wolf, M. M. Teaching conversation-related skills to pre-delinquent girls. *Journal of Applied Behavior Analysis*, 1976, **9**, 371.
- Maloney, K. B. and Hopkins, B. L. The modification of sentence-structure and its relationship to subjective judgments of creativity in writing. *Journal of Applied Behavior Analysis*, 1973, **6**, 425-433.
- Maloney, K. B., Jacobson, C. R., and Hopkins, B. L. An analysis of the effects of lectures, requests, teacher praise, and free time on the creative writing behaviors of third-grade children. In E. Ramp and G. Semb (Eds), *Behavior analysis: Areas of research and application*. Englewood Cliffs, New Jersey: Prentice-Hall, 1975.
- McAllister, L. W., Stachowiak, J. G., Baer, D. M., and Conderman, L. The application of operant conditioning techniques in a secondary school classroom. *Journal of Applied Behavior Analysis*, 1969, **2**, 227-285.
- McFall, R. M. Behavioral training: A skill acquisition approach to clinical problems. In J. T. Spence, R. C. Carson, and J. W. Thibaut (Eds), *Behavioral approaches to therapy*. Morristown, New Jersey: General Learning Press, 1976.
- Nelson, R. O. and Bowles, P. E. The best of two worlds-observation with norms. *Journal of School Psychology*, 1975, **13**, 3-9.
- Patterson, G. Interventions for boys with conduct problems: multiple settings, treatments and criteria. *Journal of Consulting and Clinical Psychology*, 1974, **42**, 471-481.
- Porterfield, J. K., Herbert-Jackson, E., and Risley, T. R. Contingent observation: an effective and acceptable procedure for reducing disruptive behavior of young children in a group setting. *Journal of Applied Behavior Analysis*, 1976, **9**, 55-64.
- Thomas, J. D., Presland, I. E., Grant, M. D., and Glynn, T. L. Natural rate of teacher approval and disapproval in grade-7 classrooms. *Journal of Applied Behavior Analysis*, 1978, **11**, 91-94.
- Van Houten, R. *Learning through feedback*. New York: Human Sciences Press, *in press*. (a)
- Van Houten, R. The performance feedback system: Generalization of effects across time. *Child Behavior Therapy*, *in press*. (b)
- Van Houten, R., Morrison, E., Jarvis, R., and McDonald, M. The effects of explicit timing and feedback on compositional response rate in elementary school children. *Journal of Applied Behavior Analysis*, 1974, **7**, 547-555.
- Walker, H. M. and Hops, H. Use of normative peer data as a standard for evaluating classroom treatment effects. *Journal of Applied Behavior Analysis*, 1976, **9**, 159-168.
- Warren, S. F., Rogers-Warren, A., and Baer, D. M. The role of offer rates in controlling sharing by young children. *Journal of Applied Behavior Analysis*, 1976, **9**, 491-497.
- White, M. A. Natural rates of teacher approval and disapproval in the classroom. *Journal of Applied Behavior Analysis*, 1975, **8**, 367-372.
- Willner, A. G., Braukmann, C. J., Kirigin, K. A., Fixsen, D. L., Phillips, E. L., and Wolf, M. M. The training and validation of youth-preferred social behaviors with child-care personnel. *Journal of Applied Behavior Analysis*, 1977, **10**, 219-230.
- Wolf, M. M. Social validity: the case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 1978, **11**, 203-214.

Received 31 July 1978.

(Final Acceptance 20 July 1979.)