

Pandora's Box Reopened? The Assessment of Social Skills

James P. Curran¹

Accepted March 23, 1979

Social skills are viewed as an organizing principle for relating a variety of superficially dissimilar behaviors under a single megaconstruct. The underlying theme of this article is that since social skills are a megaconstruct we must be constantly vigilant lest we begin to regard them as a static, invariant dispositional-trait-like unit. Within this context, various theoretical-definitional issues are discussed and reduction of these issues is urged for any particular assessment task if adequate assessment is to be achieved. Social skills are viewed as response capabilities inferred from performance. Since performance is influenced by other parameters (e.g., cognitive factors – emotional states), these other parameters may need to be measured depending on the assessment task. The advantages and disadvantages of both molar unit and molecular unit recording of social skills are discussed. It is concluded that neither type of unit should be regarded as the “proper” measurement unit. The “proper” unit size is an empirical question which may differ as a function of the assessment task. An analogy is provided regarding the assessment of the megaconstruct of intelligence. It is hoped that we can profit from an examination of the strategies employed in intellectual assessment. Behavior assessors are urged to refamiliarize themselves with various psychometric theories and strategies to increase their methodological sophistication with respect to the assessment of social skills.

KEY WORDS: social skills assessment; social-behavioral assessment; social competency; social inadequacy; assessment of assertive behavior.

¹Brown University Medical School and Veterans Administration Medical Center, Davis Park, Providence, Rhode Island 02908.

INTRODUCTION

Behavioral assessment arose, in part, because of the dissatisfaction of behaviorally oriented researchers and clinicians with the utility of the trait-dispositional assessment model for both treatment and research. Since there appear to me to exist some parallels in the assessment of social skills and the assessment of traits, I feel that it may be useful in this article to resurrect some of the historical controversies involved in the assessment of traits. Wiggins (1973, p. 368) defined a trait as "a hypothetical construct which provides an organizing principle for relating a variety of superficially dissimilar behaviors under a single dispositional unit." Mischel (1968) articulated many of the objections against the assumptions of the trait model. He noted that defining a trait as a dispositional or causal variable was basically a circular argument. Mischel (1968) further cited evidence arguing against the consistency of traits as well as their lack of cross-situational generality. Mischel (1968) rebuked trait theorists for ignoring environmental factors in their assessment and made a case for regarding observed behaviors as samples rather than as signs of criterion measures. In point of fact, most studies examining the relative contributions of persons and situational variables indicate that the contributions of both situation and individual differences are approximately equal and quite small in comparison to the variance contributed by interactions of situations, responses, and individual differences (Wiggins, 1973). Furthermore, the interpretation of behavior as either samples or signs, while providing a heuristic dichotomy, is, in most cases, nonrepresentative of actual theoretical positions, and these two positions are best thought of as end-points that define a continuum of conceptualization (Wiggins, 1973). For example, social behaviorists often employ the construct of anxiety as a unit to represent a complex pattern of physiological, behavioral, and cognitive responses, and do speak of individual differences with respect to characteristic levels of anxiety arousal (Paul & Bernstein, 1973).

Behavioral assessment was a relatively easy task when we were concerned with relatively unitary behaviors in highly constrained settings, such as "out of seat behavior" in a classroom. In those cases, we were concerned with relatively discrete behaviors which could be highly operationalized so that little inference on the part of the observer was required. Since there was no need to generalize to behaviors outside of the classroom, one could estimate in principle (i.e., by time-sampling procedures) all instances of the behavior of interest. Clearly, the assessment of social skills is not so simple, for here we are dealing with a mega-construct involving many different and complex topologies in varied and relatively unrestrained situations. Whether the term "social skills" proves useful as an organizing principle for specifying relationships among behavioral attributes which are topographically dissimilar is an empirical question.

Clearly, the conceptualization of social skills training appears to possess some utility as a treatment procedure (Curran, 1977; Hersen & Bellack, 1976).

However, the assessment procedures used to measure the success of social skills treatment have been questioned since most instruments lack a sound psychometric foundation (Curran, 1977; Curran & Mariotto, in press; Hersen & Bellack, 1977). Since the reliable and valid assessment of variables is a fundamental prerequisite to clinical research (Paul, 1969), we need to improve our methodological procedures.

We must first keep in mind that we are addressing a megaconstruct or organizing principle. We should not regard this construct as a causal variable, as did some trait theorists, but as response capabilities. The probability of occurrence of these responses in criterion situations is determined by environmental factors, person variables, and the interaction between environmental factors and individual differences. We should also bear in mind that the questions of behavioral consistency or inconsistency and cross-situational specificity or generality are empirical questions and should not be determined *a priori* on the basis of theoretical perspectives. We should be wary of falling into the same assumptive traps of previous trait theorists. We should not be blinded to the issues involved in the megaconstruct of social skills nor to the sophisticated methodology needed to assess the construct properly. The first issue to be addressed in this article is the theoretical-definitional issue, which is of utmost importance because we cannot measure well what we cannot define.

THEORETICAL-DEFINITIONAL ISSUES

The term "social skills" was chosen to appear in the title of this article after some deliberation. The plural term "skills" was used to connote the complexity of the response classes and the differences in topologies subsumed under the term. The use of the term "social skill deficit" was avoided because the objective of our treatment program is to teach all the requisite skills needed for successful performance in the criterion situations. A focus on just measuring deficits could lead us to ignore important requisite skills for any particular criterion situation. We need to measure both requisite behavior and deficit behavior, responses to be increased and responses to be eliminated and should not assume, for example, that because an individual does not exhibit any obnoxious social behavior that he behaves skillfully.

The term "social skills" was used in preference to terms such as "social competence" and "social adequacy" because of its more neutral connotations. The term "adequacy" seems to connote a minimum requisite skill (i.e., a barely sufficient quality) for a particular situation while the term "competence" seems to connote the optimal requisite skills (i.e., most effective responses) for a particular situation. Since the objectives of some studies (e.g., outcome studies with psychiatric patients) appear to emphasize teaching minimal skills, and the objectives of other studies (e.g., outcome studies with dating-anxious college

students) appear to focus on more advanced skills, the more neutral term “social skills” seems preferable. It should be noted that the response classes and processes involved in minimal and optimal performance differ; consequently, both our training procedures and assessment paradigms might also have to differ as a function of whether we are focusing on minimal or optimal skills.

Another theoretical-definitional decision that needs to be made is the nature of the response classes that are to be regarded as falling under the rubric of the megaconstruct of social skills. It seems evident that the response classes in question should be primarily social in nature. The adequacy with which an individual ties his shoes does not appear to be a fundamental concern for the assessment of social skills. However, it is difficult to determine *a priori* which acts are primarily social in nature. For example, a confrontation with an armed assailant involves a social interaction, but should the act or acts involved in successfully fleeing an assailant be regarded as an instance of social skills? We cannot exclude the act of running strictly because it is a motoric act because other primarily motoric acts such as eye contact are regarded by investigators as an instance of social skills. I also feel that the term “social skills” should be limited to behavioral acts and that we should exclude nonbehavioral components such as cognitive processes from our definition. My opinion, however, does not seem to be shared by other investigators. For example, Liberman, Vaughn, Aitchison, and Falloon (1977) include in their conceptual model of social skills the receiving, comprehension, and processing of interpersonal messages. While these cognitive processes are extremely important in social skills performance and should be measured, my own bias is that if we don't begin to put some limitations on our definition of social skills, it will be expanded to include all human behavior. I am troubled by such an expansion when we are far from a definition of social skills with respect to behavioral acts and feel that if we are to expand the megaconstruct too far, it will render the term meaningless (Curran, *in press*). As previously mentioned, it may prove impossible to decide *a priori* what behavioral acts should serve as indicators of social skills for any particular criterion situation. The importance of an act as an indicator of social skills will vary depending on situational variables; consequently, the utility of any particular act as an indicator of social skills must be determined on an empirical basis for each of our criterion situations.

The most widely used definition of social skills was proposed by Libet and Lewinsohn (1973, p. 311). They define social skills as “the complex ability to maximize the rate of positive reinforcement and to minimize the strength of punishment from others.” This definition does not exclude nonsocial behaviors, which is an issue that needs to be addressed. It also assumes that we know what is reinforcing and what is punishing for a particular subject and that we are able to anticipate not only the short-term consequences but also the long-term consequences of a response. Clearly, when we are making judgments regarding whether an act will maximize the rate of positive reinforcement and/or minimize

the strength of punishment, we are dealing with probabilities. Furthermore, in defining the skillfulness of a response solely by consequences, we are neglecting the general appropriateness of the response. Does the end justify the means? Theoretically, the consequences criterion can be applied for any particular individual in any particular situation. However, practically speaking, when clinicians train subjects in social skills they are not usually preparing them for specific idiosyncratic situations but rather attempting to teach the behaviors which are *generally* regarded as appropriate and which *generally* result in favorable consequences. Table I is presented as an illustration of this dilemma. In this table, I have artificially dichotomized the judgments regarding the social appropriateness of the act and the favorableness of the consequences. Cells 1 and 3 present no judgmental problems; in cell 1 the individual behaves in an appropriate fashion and receives favorable consequences, and in cell 3 the individual behaves inappropriately and obtains unfavorable consequences. When an individual behaves appropriately but does not receive favorable consequences (cell 2), the therapist has several options such as inducing the patient to change his potential sources of reinforcing agents, including into therapy the present reinforcing agents and attempting to modify their behavior, etc. Cell 3 presents a different type of dilemma to the clinician. Here a patient may be acting in a socially unacceptable manner but is positively reinforced for such acts. It may be very difficult for the therapist to change this particular type of patient's mode of responding because of the immediate reinforcement received for such acts. The therapist may chose to focus in on how the client's manner of acting does not generalize very well outside of his immediate source of reinforcing agents and/or focus in on the potential long-term unfavorable consequences of that manner of responding. Again, the therapist may also include the patient's present reinforcing agents and attempt to change their contingencies. Regardless, in our assessment of therapeutic change, it is impractical to deal with idiosyncratic situations and our assessment by necessity involves judgments regarding the *general* appropriateness of the response.

Table I. Social Skills Defined in Terms of Consequence and General Stylistic Appropriateness

		Consequences	
		Favorable	Unfavorable
General stylistic appropriateness	Cell 1	Behavior appropriate/ consequences favorable	Cell 2 Behavior appropriate/ consequences unfavorable
	Cell 4	Behavior inappropriate/ consequences favorable	Cell 3 Behavior inappropriate/ consequences unfavorable

In this section, I have been discussing the assessment of the megaconstruct of social skills as if it existed in a vacuum. In fact, our assessment of social the skill it is no longer readily available because of atrophy (e.g., the decay in that when we attempt to measure our subjects' social skills capabilities, we do so by measuring performance in situations and social performance is a function of numerous other factors in addition to response capabilities. In the next section, I will be discussing some of these factors which need to be considered in our assessment of social skills.

ETIOLOGICAL AND MAINTAINING FACTORS OF POOR SOCIAL PERFORMANCE

Poor social performance may be due to an actual skills deficit wherein the individual may never have learned the appropriate behavior or having learned the skill it is no longer readily available because of atrophy (e.g., the decay in skills often found in chronically institutionalized psychiatric patients). Another interpretation of poor social skills performance is that an individual, while possessing the requisite skills in his repertoire, also possesses competing and higher-order probability responses which are judged inappropriate for the situation.

Poor social performance may also be the result of an interference mechanism which inhibits and/or disrupts the effective application of social response capabilities. For example, certain emotional states such as anxiety may interfere with social performance. Anxiety may be associated with certain types of social situations through a classical conditioning process (regardless of whether the individual possesses the requisite skills) and may so interfere with skill application that the performance is judged as inadequate. Cognitive processes may also interfere with smooth skill application through various mechanisms. For example, if an individual misperceives social cues, then his inadequate performance may not be the result of an insufficient behavioral repertoire but rather due to misperception of social cues. Faulty cognitive reasoning and illogical assumptions may also lead to misinterpretations of perceived cues and ultimately to inadequate social performance. Internal cognitive standards with respect to acceptable criterion behavior may be so unduly stringent in some individuals that they become overly critical of their performance, which may lead to anxiety and subsequently to inadequate performance. As mentioned previously, some of these cognitive functions (e.g., perceptual and processing functions) are regarded by some investigators (Lieberman *et al.*, 1977) as social skills, although my own bias is to exclude them from the definition of social skills.

Table II contains a minimodel of some factors which need to be addressed when assessing poor social skills performance. This minimodel is not meant to be inclusive of all potential factors, and it is greatly simplified for presentation.

Table II. Minimodel of Social Skills Performance

		Social skills	
		Adequate	Inadequate
Interference mechanisms	Absent	Cell 1 Adequate skills/ interference absent	Cell 3 Inadequate skills/ interference absent
	Present	Cell 2 Adequate skills/ interference present	Cell 4 Inadequate skills/ interference present

For example, social skills repertoire is viewed as a dichotomous variable (i.e., adequate or inadequate) and the adequacy of the repertoire is judged for just one criterion situation. Still, this minimodel serves as an illustration of the complexities involved in assessing the nature of poor social skills performance. Three types of hypotheses regarding poor social performance are included in this minimodel: skills deficits, faulty cognitive-evaluation processes, and conditioned anxiety. Individuals in cell 1 of Table II can be characterized as possessing adequate social skills repertoires; they accurately perceive social cues, they realistically evaluate their performance as adequate, and they are not conditionally anxious to the criterion situation. Individuals in cell 2 can be characterized as also possessing an adequate behavioral repertoire, but because of one or more interference mechanisms they are not performing up to their capability in the criterion situation. Possible sources of interference could be conditioned anxiety, misperception of social cues, overly rigid standards with negative self-evaluations, etc. Individuals in cell 3 lack an adequate repertoire but do not appear to be unduly anxious in the criterion situation. These individuals may never have learned the appropriate behaviors or emit a high base rate of inappropriate behaviors which are judged as unskilled. These individuals may be insensitive to negative feedback because either they misperceive the feedback or it carries little consequence for them. Although there are no empirical data to support my contention, I believe that the majority of individuals who are judged unskilled because of their high rate of obnoxious behaviors fall into cell 3. Individuals in cell 4 can be characterized as possessing both an inadequate behavioral repertoire and a high level of anxiety in a criterion situation. These individuals accurately perceive their performance as inadequate and are sensitive to the negative consequences of their inadequate performance. These individuals can be characterized as experiencing reactive anxiety (Paul & Bernstein, 1973).

Although the evidence is somewhat circumstantial, there are data (e.g., in the dating skill area, Curran, Wallander, & Fischetti, 1977; Greenwald, 1978;

Pilkonis, 1977) which appear consistent with the classification scheme presented in this minimodel. As mentioned previously, this minimodel is not meant to be all-inclusive but is provided in order to illustrate the complex relationships among factors involved in social performance and to indicate some parameters which need to be measured in order to thoroughly assess poor social performance. Of course, the thoroughness of our assessment and the parameters to be assessed should be dictated by the purpose of the assessment. For some prediction purposes, it may not be necessary, for example, to evaluate cognitive-evaluative processes, while for other predictions it may be necessary to include factors not presented in this minimodel. In any case, assessment should be multimodal and include self-report, observational, and physiological measures whenever appropriate because of the different sources of error associated with the various modes and because of the lack of convergence within the organism itself (Curran, 1977).

Next, I would like to address some issues and problems involved with the appropriateness of the behavioral units selected for our measurement of social skills. I have artificially dichotomized the unit size question into molar vs. molecular, although unit size is actually a gradient. Assessment should be viewed as a multilayered process along this gradient. The "proper" unit size should not be determined on theoretical grounds but should be determined on an empirical basis. The choice of unit size is dependent on the assessment questions asked and the empirical relationships established. The "proper" units of observation are those which have been empirically determined as possessing the highest degree of criterion relevance, and these units may vary depending on various assessment questions. Behaviorally oriented assessors have generally preferred molecular units because they usually can be operationalized with greater precision and consequently require less inference on the part of the observers. However, as Wiggins (1973, p. 325) has noted, "the relationship between unit size and criterion relevance cannot be stated dogmatically. Narrowly-defined behavioral attributes run the risk of a high degree of specificity, which may preclude the possibility of generalizability to criterion behaviors. On the other hand, units that are too global in nature may yield only vacuous statements that are true of everyone or of no one."

ASSESSMENT AT THE MOLAR LEVEL

Examples of molar level assessment of social skills are numerous in the literature. Investigators often ask the subjects themselves, trained observers, untrained significant others, etc., to make global ratings of subjects' social skills level, usually on some sort of Likert-type scale. In many cases, these raters are given a definition of social skills and/or training in making judgments regarding social skill levels. However, one can find many cases in the literature

where no definition or training is given to the raters (this is especially true for self-report ratings). Even in those cases where definitions are given, these definitions lack a good deal of precision; hence raters' evaluations are quite dependent on their own inferential processes.

Requesting raters to make molar judgments regarding the imprecisely defined construct of social skills creates numerous methodological problems. Individual and cultural differences exist regarding the criteria used in evaluating the appropriateness of a social act for a particular social situation, and, because these molar judgments are so dependent on the inferential process of the observer, one needs to be concerned about both reliability and ecological validity issues. The selection of raters in itself is a vexing problem (Curran, in press). One strategy proposed by Curran (in press) is to select raters who understand the value system of the significant others involved in the criterion situation. Since these significant others are crucial with respect to dispensing reinforcement contingent on the social skills performance of the subjects, it would be important for the raters to understand their value system. For example, in treating dating-anxious college students, a logical choice for raters would be opposite-sex peers of approximately the same age and educational level. In other cases, the decision regarding choice of observers is more difficult. For example, in training juvenile delinquents, which group of individuals should serve as raters: other delinquents, nondelinquent same-sex peers, parents of the juvenile offenders, parole officers, etc.?

When training observers to evaluate a subject's overall social skills, it is important that these raters possess some communality with respect to the theoretical-definitional issues discussed earlier in this article. For example, it must be clear to the raters whether they are rating minimal or optimal behaviors. Raters should be given training to familiarize themselves with the anchor points of the ratings scale as well as midrange points of the scale. This is generally done when investigators employ raters from their research team, sometimes done when using significant others as raters (e.g., nurses, relatives), but rarely done when asking the subject to rate himself or herself. Asking untrained, naive subjects to rate themselves on a nebulous and complex construct such as social skill will, in all probability, result in large differences in the rating process due to subjects' own inferential interpretation of the theoretical-definitional construct. It is no wonder that investigators (Farrell, Mariotto, Conger, Curran, & Wallander, 1979; Curran, Monti, Corriveau, Hay, Hagerman, & Zwick, 1978) have found a lack of correspondence between self-report and others' ratings of social skill.

Even when raters have been trained, investigators need to be concerned with the problem of rater "drift." That is, if raters were asked to make global ratings of social skill over a period of time, their judgments regarding the construct may change and drift from the criteria established by the investigator. Reliability among observers may remain high, but agreement with the criterion may decrease (DeMaster, Reid, & Twentyman, 1977).

Although molar recordings are fraught with many difficulties, they are likely to be continually employed. We are not at this time (or in the immediately foreseeable future) able to specify for any of our criterion situations with any degree of precision the most important components of social skills: how these components interface in a complex interaction, how these components should be differentially weighed in making a judgment, etc. In addition, we have very little data (and the data we do have are disappointing, Curran & Mariotto, in press) regarding the degree of generalizability of behavior observed in our laboratory-based assessment to behavior observed in a naturalistic criterion situation. Since investigators are ultimately interested in subjects' actual behavior in the criterion situation, and since molecular recordings in naturalistic situations are difficult to obtain, we will probably see a continual reliance on molar measurement in the social skills area. A second and related issue arguing for molar recordings lies in our concerns regarding the adequacy of our assessment sampling procedures. An individual's social skills level differs across occasions, settings, etc., and because investigators may be interested in a subject's social skills performance across many facets, they will continue to rely on molar recordings because of the ease in which they are obtained. Another reason for the use of molar recordings is that they often resemble the types of judgments made in the criterion situations. For example, when we conduct social skills training with psychiatric patients, the ultimate goal of such treatment is that it will enable these patients to more successfully cope with problematic social situations. Those significant others in the natural environment evaluating the adequacy of the subjects' responses are making unsystematic molar judgments. An argument could be made that molar assessment strategies may have greater generalizability to these criterion ratings because of the similarity which exists between them, but, of course, this is an empirical question.

ASSESSMENT AT THE MOLECULAR LEVEL

The major advantage of molecular unit recording over molar unit recording is the degree of precision in which molecular units can be defined with the usual subsequent increase in generalizability of recording across observers. The major problem lies in determining those molecular units which are criterion relevant. It does us little good to be able to reliably measure a particular unit of behavior if it is not related to the behavior which we wish to predict. The strategies employed in the social skill literature to decide which molecular units to record have varied considerably. Two illustrations will be provided as examples. One instrument frequently used in the social skills literature which involves molecular unit recording is the Behavioral Assertiveness Test - Revised (BAT-R) (Eisler, Hersen, Miller, & Blanchard, 1975). The BAT-R consists of 32 situations presented on audio tape. The subjects' responses to these test

items are videotaped and are rated for a number of molecular units such as duration of looking, duration of reply, and latency of response. Other behaviors rated on the BAT-R appear to be somewhat more molar such as affect, compliance, content requesting new behavior, and a global rating of overall assertiveness. The molecular units scored on the BAT-R were generated by several experienced clinicians who were asked to list specific behaviors that they felt might be related to judgments of assertiveness. Recent evidence (Bellack, Hersen, & Turner, 1978; Bellack, Hersen, & Lamparski, in press) would seem to indicate that the units of behavior recorded on the BAT-R do not generalize well to more naturalistic situations, although methodological flaws make it difficult to draw conclusions from these two studies (Curran, 1978).

Another example of an attempt to delimit microunits of social skills is illustrated by the work of Conger, Wallander, Ward, and Mariotto (1978). Conger *et al.* presented stimulus tapes consisting of males differing in heterosexual social skills to 62 male and 73 female undergraduates. The undergraduate judges were asked to make global ratings of skill and to list those behaviors which they used as cues in making these global ratings. The cues generated were then analyzed for content and combined into a hierarchical classification system by the investigators. At the lowest level of the hierarchical classification system were molecular units such as smiles, with each higher order of the hierarchy containing more complex units. The judges were successful in differentiating the skill level of the subjects presented on tapes. Conger *et al.* are now in the process of determining whether the cues generated by the judges were actually representative of the behavior of the subjects on the tapes. If the nominated cues do match the subjects' behaviors, then Conger *et al.* must empirically determine whether, indeed, these units were the major cues utilized in discriminating social skill level. Of course, the utility of these molecular units in predicting assessment questions needs to be determined.

Molecular unit recording requires a good deal of effort both in delimiting those units which are relevant and in the actual recording of those units. Certainly, more effort is required than in making molar unit recordings. For this reason, investigators generally record molecular units from a small sample of a subject's behavior in rather constrained settings. For example, on the BAT-R, a narrator describes a particular situation to a subject and a confederate issues a prompt after which the subject gives a brief, often one-sentence reply. The representativeness of such assessment situations to more naturalistic interactions is questionable (Bellack *et al.*, 1978; Curran, 1978). In addition, molecular unit recording often ignores important parameters such as the timing and sequencing of behaviors (Fischetti, Curran, & Wessberg, 1977).

Molecular unit analysis is appealing because of the precision with which the behaviors of interest can be defined and the relative lack of inference required on the part of the observer. However, molecular unit recording requires a great deal of effort both in determining the relevancy of such behaviors and in actually

recording them. The difficulty involved in molecular unit recording should not discourage us from such endeavors, because they are necessary if we are to fully comprehend the dimensions of social skills. The utility of such molecular unit recording in answering different types of assessment questions must, of course, be determined on an empirical basis.

ASSESSMENT AT INTERMEDIATE LEVELS

As mentioned previously, I have artificially dichotomized the unit size selected for observation into molar vs. molecular behaviors. In actuality, unit size differs along a continuum and many examples of intermediate level assessment can be found in the social skills literature. For example, Lowe and Cautela (1978) have developed a questionnaire purportedly measuring social skills which can be filled out by the subjects themselves and/or significant others. The items on the scale include questions such as whether the subject easily becomes angry (not at all to very much). These units require less inference than an overall social skills rating but are more inferential than many types of molecular units. Of course, whether these behavioral units are good indicators of social skills and whether they are related to criterion behavior must be empirically established.

An interesting intermediate assessment level strategy employed by Goldsmith and McFall (1975) and MacDonald (1978) appears to be something of a hybrid; that is, raters are requested to make overall ratings regarding the social appropriateness of a particular response but are also provided with a scoring key to assist in making this decision. These investigators employed criterion keying procedures wherein they first elicited numerous responses to their assessment items from the population in question. This was followed by a response evaluation phase wherein a sample of significant others in the environment, who would typically label the behavior patterns as being effective or maladaptive, were asked to judge the appropriateness of each response generated in the response enumeration phase. The advantages of a scoring key are obvious; however, it should be kept in mind these examples of criterion keying were conducted on contrived, relatively brief laboratory assessment situations and that the utility of such types of assessment in predicting criterion behavior must be demonstrated.

CONCLUSION

It occurred to me when writing this article that the problems and issues involved in the assessment of the megaconstruct of social skills were hauntingly reminiscent of the issues and problems in assessing another construct in psychology, namely that of intelligence. Intelligence, like social skills, is an organizing

principle for relating a variety of superficially dissimilar behaviors. Binet did not regard his measure of intelligence as a static dispositional trait but as a collection of abilities. These abilities were measured as response capabilities, and it was understood that subjects' performances were also influenced by nonintellectual factors. These abilities consisted of numerous verbal and nonverbal factors which Binet attempted to sample (albeit somewhat nonempirically) in order to predict the criterion behaviors (i.e., school performance). The number, type, and organization of these abilities are still a subject of great controversy in psychology (Maloney & Ward, 1976). Vernon (1950) proposed a hierarchical model of intelligence (somewhat analogous to Conger *et al.*, 1978, hierarchical structure of heterosexual-social skills) as a resolution to the abilities controversy. In Vernon's model, a general factor (similar to Spearman's *g* factor) is placed at the top of the hierarchy, followed by two major group factors (verbal, numerical, educational, and practical, mechanical, spatial, physical) and then multiple minor group factors (similar to Thurston's primary factors) and then numerous specific factors (similar to Guilford's specific factors). As mentioned previously, the optimal level within any hierarchical structure for prediction may vary depending on the assessment task, although parenthetically it should be noted that for intelligence the use of specific factors has not resulted in any significant increase in prediction over the use of *g* for most prediction tasks (McNemar, 1964).

Numerous parallels exist between the constructs of social skills and intelligence. Intellectual performance is somewhat consistent over time, and what few data we have with respect to social skills performance indicate some consistency (Trower, Bryant, & Argyle, 1977). Intellectual performance is influenced by situational variables, and the existing data (Curran & Mariotto, in press) with respect to social skills indicate that skills performance is greatly influenced by situational variables. Theoretical-definitional issues are far from being resolved for either construct. The judgments involved in evaluating intellectual and social skills performances are influenced by social and cultural factors which differ among individuals (witness the controversy over the cultural fairness of intelligence tests).

Perhaps because of the parallels associated with the assessment of the constructs of social skills and intelligence, we may be able to profit from an examination of the strategies employed in the assessment of intelligence that have evolved over the years. The major problem in the social skills area is the "criterion problem." Our criterion measures are imperfect just as Binet's criterion of teachers' ratings was imperfect. Binet developed a "bootstrap" approach to the criterion problem, and we are beginning to witness a similar approach in the social skills area (Curran & Mariotto, in press). Binet selected items for his test on an empirical basis; e.g., he selected items which differentiated older from younger children, which had good internal consistency, etc. Parenthetically, it should be noted that the types of items composing Binet's scale differ in content for younger as opposed to older children. Analogously, Cox, Gunn, and Cox (1976) noted that this might be the case in the social skills assessment area.

Binet had to decide on the level of the response units for his items. He rejected the approach of Galton, who had incorrectly assumed that the ability to make fine sensory discriminations was related to intelligence. Binet instead focused on a variety of higher-order mental processes which predicted criterion behaviors better than sensory discrimination tasks. Although it is an empirical question, it might be more profitable in social skills assessment to focus on higher-order skill functions such as the ability to handle criticism than to focus on more molecular units such as eye contact if we hope to establish generalizability to criterion situations. Binet and other developers of intellectual tests strove to obtain objective scoring keys; however, as we well know, most tests of intelligence require some inference on the parts of the examiner. Likewise, recent attempts (e.g., Freedman, Rosenthal, Donahoe, Schlundt, & McFall, 1979; Goldsmith & McFall, 1975; McDonald, 1978) to provide objective scoring keys in the social skills assessment area still require inference on the part of raters. Although molar unit recording will be somewhat inferential, we must strive to objectify our scoring keys as much as possible in order to promote reliability of scoring. Another characteristic of intelligence tests that may prove useful in the measurement of social skills is the conception of a gradient of items varying in difficulty level. In most cases, we have no idea of the difficulty level of the items (e.g., the role-play of a simulated social situation) we present to our subjects in the assessment of social skills. Depending on our population, we may have problems with ceiling and floor effects which could severely interfere with our prediction of criterion behaviors.

One final point should be noted regarding the present analogy. Binet never intended that his scale be regarded as a measure of all the abilities connoted by the term "intelligence," nor did he attempt to predict all possible criterion situations. His concern was not theoretical but practical. Binet's aim was to develop a fairly objective procedure for screening out the mentally retarded from regular public school instruction. Intelligence tests do a reasonably good job in accomplishing the purpose for which they were created, i.e., the prediction of academic success, but do less well (Maloney & Ward, 1976) in predicting other possible criterion measures of intelligence (e.g., occupational success). It is highly unlikely that we will be able to develop a measure of social skills that would predict reasonably well the multitude of possible criterion situations. In all probability, we will need different assessment instruments and methodologies to predict different types of criterion measures of social skills.

In conclusion, I have noted that we are far from closure regarding the theoretical-definitional issues involved in the assessment of social skills. It is important that we begin to reach consensus with respect to these theoretical-definitional issues because we cannot measure well what we cannot define. We must also realize that in assessing social skills what we are observing is performance, and that the performance of an individual in a particular situation is dependent on many situational and person-by-situational variables. Depending

on the measurement task at hand, we may need to be assessing other constructs such as anxiety, cognitive processes, and behavioral intention. We have also seen that neither molar unit recordings nor molecular unit recordings are the "proper" units of observation for social skills; rather, the "proper" unit is determined by its degree of criterion relevance and the "proper" unit may differ depending on the assessment task.

Most importantly, it is hoped that the reader realizes that in attempting to measure a megaconstruct such as social skills we are crossing ground where previous behavioral assessors "feared to tread." When behavioral assessors were concerned with simple unitary behaviors in highly constrained settings, it was possible to stringently operationalize these behaviors and observe at least in principle all instances of these behaviors. In assessing social skills, we are attempting to assess more complex sets of behaviors which are topographically dissimilar in less constrained settings. We can no longer depend on simple measurement procedures such as counting all instances of the response class because the complexity of the construct of social skills involves such variables as moderation, timing, and sequencing. Our knowledge of traditional psychometric techniques must increase as well as our degree of methodological sophistication.

One final warning: we should be very wary of falling into the trap that previous trait theorists did when attempting to measure other psychological constructs. Social skills should not be viewed as an invariant trait. An individual's social skills performance in one particular situation may not be predictive of his performance in other criterion situations. Social skills are not a disposition but a response capability. Questions of stability and cross-situational generality need to be determined on empirical bases. The latch on Pandora's box has been unfastened. We must be cautious lest someone is foolish enough to open the damn box.

ACKNOWLEDGMENT

The author wishes to thank Dr. Peter M. Monti for his valuable comments on an earlier draft of this article.

REFERENCES

- Bellack, A. S., Hersen, M., & Turner, S. M. Role-play tests for assessing social skills: Are they valid? *Behavior Therapy*, 1978, 9, 448-461.
- Bellack, A. S., Hersen, M., & Lamparski, D. Role-play tests for assessing social skills: Are they valid? Are they useful? *Journal of Consulting and Clinical Psychology*, 1979, 47, 335-342.
- Conger, A. J., Wallander, J., Ward, D., & Mariotto, M. J. Peer judgments of heterosexual-social anxiety and skill: What do they pay attention to anyhow? Paper presented at Association for Advancement of Behavior Therapy, Chicago, 1978.

- Cox, R. D., Gunn, W. B., & Cox, M. J. A film assessment and comparison of the social skillfulness of behavior problems and non-problem male children. Paper presented at the Association for the Advancement of Behavior Therapy, New York, 1976.
- Curran, J. P. Skills training as an approach to the treatment of heterosexual-social anxiety: A review. *Psychological Bulletin*, 1977, 84, 140-157.
- Curran, J. P. Comment on Bellack, Hersen, & Turner's paper on the validity of role-play tests. *Behavior Therapy*, 1978, 9, 462-468.
- Curran, J. P. Social skills: Methodological issues and future directions. In A. S. Bellack & M. Hersen (Eds.), *Research and practice in social skills training*. New York: Plenum Press, in press.
- Curran, J. P., & Mariotto, M. J. A conceptual structure for the assessment of social skills. In M. Hersen, R. Eisler, & P. Miller (Eds.), *Progress in behavior modification*. New York: Academic Press, in press.
- Curran, J. P., Wallander, J., & Fischetti, M. The role of behavioral and cognitive factors in the maintenance of heterosexual-social anxiety. Paper presented at the Midwestern Psychological Convention, Chicago, 1977.
- Curran, J. P., Monti, P. M., Corriveau, D. P., Hay, L. R., Hagerman, S., & Zwick, W. R. The generalizability of procedures for assessing social skills and social anxiety in a psychiatric population. Unpublished manuscript, Brown University Medical School, 1978.
- DeMaster, B., Reid, J., & Twentyman, C. The effects of different amounts of feedback in observers' reliability. *Behavior Therapy*, 1977, 8, 317-329.
- Eisler, R. M., Hersen, M., Miller, P. M., & Blanchard, E. B. Situational determinants of assertive behaviors. *Journal of Consulting and Clinical Psychology*, 1975, 43, 330-340.
- Farrell, A. D., Mariotto, M. J., Conger, A. J., Curran, J. P., & Wallander, J. L. Self- and judges' ratings of heterosexual-social anxiety and skill: A generalizability study. *Journal of Consulting and Clinical Psychology*, 1979, 47, 164-175.
- Fischetti, M., Curran, J. P., & Wessberg, H. W. Sense of timing: A skill deficit in heterosexual-socially anxious males. *Behavior Modification*, 1977, 1, 179-193.
- Freedman, B. J., Rosenthal, L., Donahoe, C. P., Schlundt, D. G., & McFall, R. M. A social-behavioral analysis of skill deficits in delinquent and nondelinquent adolescent boys. *Journal of Consulting and Clinical Psychology*, 1979, 46, 1148-1163.
- Goldsmith, J. B., & McFall, R. M. Development and evaluation of an interpersonal skill-training program for psychiatric patients. *Journal of Abnormal Psychology*, 1975, 84, 51-58.
- Greenwald, D. P. Self-report assessment in high- and low-dating college women. *Behavior Therapy*, 1978, 9, 297-299.
- Hersen, M., & Bellack, A. S. Social skills training for chronic psychiatric patients: Rationale, research findings, and future directions. *Comprehensive Psychiatry*, 1976, 17, 559-580.
- Hersen, M., & Bellack, A. S. Assessment of social skills. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Handbook for behavioral assessment*. New York: Wiley, 1977.
- Lieberman, R. P., Vaughn, C., Aitchison, P. A., & Falloon, I. Social skills training for relapsing schizophrenics. Funded grant from the National Institute of Mental Health, 1977.
- Libet, J. M., & Lewinsohn, P. M. Concept of social skills with special reference to the behavior of depressed persons. *Journal of Consulting and Clinical Psychology*, 1973, 40, 304-312.
- Lowe, M. R., & Cautela, J. R. A self-report measure of social skill. *Behavior Therapy*, 1978, 9, 535-544.
- MacDonald, M. L. Measuring assertion: A model and method. *Behavior Therapy*, 1978, 9, 889-899.
- Maloney, M. P., & Ward, M. P. *Psychological assessment: A conceptual approach*. New York: Oxford University Press, 1976.
- McNemar, Q., Lost: our intelligence? why? *American Psychologist*, 1964, 9, 871-882.

- Mischel, W. *Personality or assessment*. New York: Wiley, 1968.
- Paul, G. L. Behavior modification research: Design and tactics. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status*. New York: McGraw-Hill, 1969.
- Paul, G. L., & Bernstein, D. A. *Anxiety and clinical problems: Systematic desensitization and related techniques*. Morristown, N.J.: General Learning Co., 1973.
- Pilkonis, P. A. Shyness public and private and its relationship to other measures of social behavior. *Journal of Personality*, 1977, 45, 585-595.
- Trower, P., Bryant, B., & Argyle, M. *Social skills and mental health*. Pittsburgh: University of Pittsburgh Press, 1978.
- Vernon, P. E. *The structure of human abilities*. London: Methuen, 1950.
- Wiggins, J. S. *Personality and prediction: Principles of personality assessment*. Reading, Mass.: Addison-Wesley Publishing Co., 1973.