

2001

Analogue Observational Methods in the Assessment of Social Functioning in Adults

Peter J. Norton

University of Nebraska-Lincoln

Debra A. Hope

University of Nebraska-Lincoln, dhope1@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/psychfacpub>



Part of the [Psychology Commons](#)

Norton, Peter J. and Hope, Debra A., "Analogue Observational Methods in the Assessment of Social Functioning in Adults" (2001).
Faculty Publications, Department of Psychology. 912.
<https://digitalcommons.unl.edu/psychfacpub/912>

This Article is brought to you for free and open access by the Psychology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, Department of Psychology by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Published in *Psychological Assessment* 13:1 (2001), pp. 59–72; doi: 10.1037//1040-3590.13.1.59
Copyright © 2001 by the American Psychological Association. Used by permission.
Submitted May 18, 2000; revised July 17, 2000; accepted July 24, 2000.

Analogue Observational Methods in the Assessment of Social Functioning in Adults

Peter J. Norton and Debra A. Hope

Department of Psychology, University of Nebraska–Lincoln, Lincoln, Nebraska, USA

Corresponding author – Debra A. Hope, Department of Psychology, University of Nebraska–Lincoln, Lincoln, Nebraska 68588-0308, email dhope@unl.edu

Abstract

This article provides a clinically oriented overview of analogue observational methods used in the assessment of problematic social functioning, specifically skill deficits and social anxiety. This article emphasizes role-play assessment methods, the predominant method used in clinical settings. An examination of the psychometric characteristics of analogue assessment methods is presented, followed by a review of procedural and structural considerations that may impact the quality of assessment data. Of special concern are the potential impacts of instructional variables, structured versus ideographic role-played situations, confederate characteristics and behavior, molar and molecular levels of assessment, self-ratings versus clinician ratings of functioning, and physical attractiveness. Finally, published and empirically evaluated analogue observation tests are critically reviewed with an emphasis on features that may impact their utility in clinical practice.

Dysfunctional or ineffective social functioning is a common problem seen in clinical settings. Indeed, impaired social performance is a recognized characteristic of numerous psychological disorders. For example, anxiety in, and avoidance of social situations is the hallmark of social anxiety disorder, the third most prevalent psychological disorder (Kessler et al., 1994). Social functioning deficits are also seen in schizophrenia, major depressive disorder, marital distress, and aggression. The prevalence of difficulties in social functioning makes them a common target of intervention. Therefore, it is imperative that clinicians have available valid and reliable tools for adequately assessing social behavior.

Adequate assessment of dysfunctional social behavior, however, presents unique difficulties in the clinical setting. Social behavior is always embedded in a social context. The

unique influences and social conventions of a clinical setting may mean that a client's interpersonal behavior with a clinician is not representative of his or her behavior across situations. Therefore, assessment methods that allow the clinician to access samples of social behavior in important situations from the client's daily life are needed.

Role-play assessment methods have considerable potential in clinical settings, either in lieu of, or as an adjunct to, self-report and interview measures. Although all assessment methods are susceptible to bias and error, self-report of social behavior presents special difficulties such as demand characteristics and biases, distorted recall, limited awareness of social difficulties (see Meier & Hope, 1998), lack of situational specificity (McFall, 1982), or reading ability (Andrasik, Heimberg, Edlund, & Blankenberg, 1981).

In this article we examine analogue assessment of social functioning as it could be used in nonresearch clinical settings. Given space limitations, the focus will be on deficits in social skill or social anxiety that impact one-on-one adult interpersonal behavior such as dating, assertiveness, and conversations. Communication among family members and social skills in children and adolescents pose important challenges in clinical assessment that are beyond the scope of this article. Brief construct definitions are provided, followed by a review of the psychometric characteristics of role-play methods. Then various conceptual and methodological issues facing clinicians in analogue assessment will be delineated, concluding with a review of select published role-play methods.

Social Skill and Social Anxiety

Two of the most frequently assessed problematic social behaviors include social skill deficits and social anxiety. The global concept of social skill is well recognized and improvement in social skill is frequently an important focus of assessment and treatment, yet no consistent, comprehensive definition of social skills has been identified (Meier & Hope, 1998). Trower, Bryant, Argyle, and Marziller (1978) viewed social behavior as a means to satisfy personal goals and motivations. Therefore, social skill is seen as comprising the ability to perceive interpersonal or social cues, integrate these cues with current motivations, generate responses, and enact responses that will satisfy the motives and goals. Liberman (1982) suggested that social skills can be classified as either *instrumental*, those skills necessary for gaining information or services required to fulfill needs, or *social-emotional*, skills used to initiate or maintain social relationships. Liberman (1982) also noted that social skills can be viewed from a *topographical* viewpoint, focusing on molecular behaviors such as gaze or verbal content; a *functional* viewpoint, understanding social skills in respect to outcomes of an interaction; or an *information processing* viewpoint, emphasizing the individual's ability to receive information, generate and select responses, and execute responses. Social anxiety, on the other hand, is defined as a fear of negative evaluation by others and low self-confidence when performing or interacting in social situations. Social anxiety is often accompanied by avoidance of social situations (American Psychiatric Association, 1994).

Assessment of social skill and anxiety serves three purposes. First, the assessment may play a role in diagnosis and case formulation such as the determination of whether deficits in social performance are attributable to anxiety that interferes with implementation of

social skills that the person has in his or her repertoire or whether the client's repertoire is lacking crucial social skills. Such differential formulation of the nature of social difficulties has implications for treatment planning (e.g., anxiety reduction strategies vs. social skills training). Second, detailed assessment may inform treatment by identifying relevant situational factors, the degree of social anxiety or skill deficit, specific skill deficits, or specific symptoms of anxiety. In most established treatments for social anxiety and the training of social skills, attention to specific characteristics, deficits, and situational factors are of paramount importance. Finally, ongoing assessment provides the clinician with data to monitor client progress and outcome.

Correspondence between Analogue and Naturalistic Observational Assessment

Ideally, assessment of problematic social behavior would occur through naturalistic observation. Naturalistic observation would allow the clinician or trained observer to assess social behavior within normal situational contexts and under naturally occurring reinforcers and consequences. However, naturalistic observation can involve considerable time (extra-session observations) and expense (recording devices, trained observers). Client reactivity to the presence of the observers may decrease the validity of the observation, but conducting observations without the consent of the client and significant others treads on ethical boundaries. Given these difficulties associated with naturalistic observation methods, it is not surprising that they are seldom used in clinical and research settings for assessing social behavior in adults. Consequently, clinicians and researchers more commonly employ situational analogue assessment methods (Haynes & O'Brien, 2000), primarily via observation of role-played scenarios. Role-played scenarios involve the simulation of an interaction between the client and another individual or a group in the clinical setting. Most commonly, clients are instructed to behave as they typically would and are asked to engage in one or more social interactions. For example, a client may be instructed to pretend that he or she had just been introduced to somebody at a party and to try to get to know the other person better. The subsequent conversation may last up to 10 min. Although in most cases the analogue situation is designed to approximate normal social conditions, the most basic question for the validity of analogue assessment is "Does the behavior exhibited by role-plays correspond to behavior observed in more naturalistic situations?"

As shown in Table 1, several studies have reported moderate to excellent correspondence of measures of social skill (Kern, 1991; St. Lawrence, Kirksey, & Moore, 1983; Wessberg, Marriotto, Conger, Farrell, & Conger, 1979) and anxiety (Wessberg et al., 1979) obtained from nonclinical samples in role-play and *in vivo* interactions. Notably, Wessberg et al. (1979) reported that judges' ratings of the skill and anxiety displayed by college students during role-play dating interaction were correlated with ratings of skill and anxiety obtained during a contrived waiting-room interaction with an opposite-sex confederate ($r^2 = .46$ and $.38$, respectively). Also, St. Lawrence et al. (1983), comparing the behavior of female college students in assertion-eliciting interactions, concluded that "subjects' behavior was consistent whether they role played the situation or believed it was actually occurring" (p. 32). In contrast, Gorecki, Dickson, Anderson, and Jones (1981) found that

assertiveness ratings of college students were significantly higher during role-played interactions than *in vivo* interactions.

Table 1. Summary of Selected Analogue Observation Methods for Social Behavior Assessment

Test	Citation	Study population	Social behaviors assessed	Demonstration of reliability	Demonstration of validity
Behavioral Assertiveness Test	Eisler, Miller, & Hersen (1973)	Male psychiatric inpatients	Assertiveness with a woman confederate	Interrater reliability 99–100% Agreement $r^2 = .92-.98$	Convergent High/low overall assertiveness differed in response latency, loudness, compliance, requests for new behavior, and affect. High/low overall assertiveness differed on self-reported assertiveness, but not on self-reported personality
Behavioral Assertiveness Test–Revised	Eisler, Hersen, Miller & Blanchard (1975)	Adult male psychiatric patients	Positive and negative assertiveness with women portrayed as familiar or unfamiliar	Interrater reliability > 95% Agreement $r^2 > .88$	Convergent High/low overall assertiveness differentiated on most molecular ratings and self-reported assertiveness High/low overall assertiveness equal on gaze duration, speech latency, and verbal demonstration of appreciation
Behavioral Assertiveness Test–Revised	Baggs & Spence (1990)	Low assertion women	—	Interrater reliability $r^2 = .40-.83$ (eye contact, response duration, latency, compliance, and overall assertiveness) $r^2 < .36$ for all other ratings	Discriminative Participants receiving assertiveness training improved in response latency, compliance, and overall assertiveness. No change for wait-list controls.
Behavioral Assertiveness Test–Revised (Modified)	Bellack, Hersen, & Turner (1979)	Adult psychiatric inpatients and outpatients	Behavioral and global measures of skill/assertion	Interrater reliability 87–100% agreement $r^2 = .67-1.00$	External Speech latency, eye contact, and compliance, but not smiles, praise, requests, and overall assertiveness, moderately related between role-play and naturalistic interaction Convergent Compliance and overall assertiveness, but not praise or requests, correlated between RP and structured interview

Test	Citation	Study population	Social behaviors assessed	Demonstration of reliability	Demonstration of validity
Assessment of Interpersonal Problem-Solving Skills	Donahoe et al. (1990)	Schizophrenic outpatient & nonclinical men	Social skill description, solution generation and enactment in interpersonal problems	Interrater reliability κ (General) = .76–.88 No mean difference between observer ratings (Specific) r^2 = .90–1.00 (General) r^2 = .77–1.00 Test-Retest Reliability (Specific) r^2 = .31–.71 (General) r^2 = .21–.59 Internal Consistency α (Specific) = .69–.93 α (General) = .57–.74	Discriminative Nonclinical superior to clinical sample on all general scoring scales Convergent IQ correlated with Sending Skills
Simulated Social Interaction Test	Curan et al. (1980), Curran (1982)	Male psychiatric inpatients and outpatients & male nonclinical military personnel	Social skill and anxiety	Interrater reliability 1–4% of variability attributable to differences among judges. G (judges) = .94–.98 Other G (situations) = .76–.97	Discriminative Nonclinical superior to clinical sample on social skill and anxiety External Skill ratings correlated with naturalistic observation ratings Convergent Skill ratings correlated with interview skill ratings
Social Skill Behavioral Assessment System	Caballo & Buela (1988)	College students	Skill and anxiety in opposite-sex interactions	Interrater reliability 95–99% Agreement	Convergent Observer and participant ratings of global skill correlated with molecular ratings

Test	Citation	Study population	Social behaviors assessed	Demonstration of reliability	Demonstration of validity
Ideographic Role-Play (IRP)	Kern (1991)	High, medium, and low assertion college students	Global assertion	Interrater reliability $\kappa = .84$ $r^2 = .88$ 87% Agreement Situations $r^2 = .21-.76$ (.16-.74 corrected) Overall $r^2 = .81$ (.76 corrected) Internal Consistency Situations, $\alpha = .42-.65$ Overall, $\alpha = .77$ Retest Situations, $\alpha = .20-.63$	Convergent IRP assertiveness ratings moderately related to self-report assertiveness ($r^2 = .17$ to .18) External Specific situation scores, but not total score, were moderately to strongly related to number of requests ($r^2 = .30$) and refusal quality ($r^2 = .49$) during an <i>in vivo</i> telephone conversation. Divergent Unrelated (Test $r^2 = .01$ and .01) to social desirability at Time 1 and 2
Social Situations Test	Merluzzi & Biever (1987)	High anxiety/low skill, medium anxiety and skill, and low anxiety/high skill college men	Global ratings of skill and anxiety	Interrater reliability $r^2 = .86-.92$	Discriminative High anxiety/low skill, medium anxiety/skill, and low anxiety/high skill participants generally differentiated on confederate and judge anxiety and skill ratings Convergent Confederate and judge ratings of anxiety and skill moderately ($r^2 = .06-.24$) related Divergent Anxiety and skill ratings unrelated to attractiveness rating
Disability and Assertiveness Role-Play Test	Glueckauf & Quittner (1992)	Physically disabled adults	General assertion skills	Interrater reliability $r^2 = .30-.98$ $\kappa = .68-.77$	Discriminative Assertiveness training group increased assertive responding, directed looking, and appropriate affect, and decreased passive responding, no change for control group. Return to baseline assertiveness at 6 mos.

Test	Citation	Study population	Social behaviors assessed	Demonstration of reliability	Demonstration of validity
Social Anxiety and Skill Index-B	Wessberg et al. (1979)	High, medium, and low dating frequency college men	Overall anxiety and social skill	Interrater reliability No mean difference in judge ratings on all but 1 situation Anxiety $r^2 = .53-.61$ Skill $r^2 = .52-.76$	Discriminative High frequency daters less anxious and more skillful than low frequency daters Divergent Anxiety and skill ratings unrelated to attractiveness ratings External Anxiety and skill in waiting-room and role-plays moderately related ($r^2 = .12$ and $.14$) Skill better in role-play than waiting-room interaction
Study-Specific Unstructured same-sex interaction	Alden & Bieling (1998)	Socially anxious and nonsocially anxious college women	Likability, intimacy, and appropriateness	Interrater reliability Intimacy $r^2 = .86$ Average likability $r^2 = .59$ Average appropriate $r^2 = .90$	Discriminative In negative appraisal condition, anxious participants rated less likeable and appropriate than nonanxious participants. No differences in positive appraisal condition.
Study-Specific Structured opposite-sex interactions	Bellack, Hersen, & Lamparski (1979)	College students	Behavioral indices and overall rating of skill	Interrater reliability $r^2 = .48-.96$ 62% Agreement	External Analogue and naturalistic moderately valid for women (36% of relationships statistically significant), minimally valid for men (22% of relationships statistically significant)
Study-Specific Same- and opposite-sex unstructured interaction and public speech	Beidel, Turner, & Dancu (1985)	Socially anxious and nonanxious adults	Behavioral indices and overall rating of social skill and anxiety	Interrater reliability $r^2 = .67-.88$	Discriminative Socially anxious participants rated as overall somewhat more anxious and less skilled than nonanxious participants. No differences on any behavioral indices between socially anxious and non-anxious men and women, except socially anxious men tended to be rated lower on appropriateness of gaze

Test	Citation	Study population	Social behaviors assessed	Demonstration of reliability	Demonstration of validity
Study-Specific Standardized and personalized interactions	Chiauzzi, Heimberg, Becker, & Gansler (1985)	Mood disorder outpatient adults	Positive and negative assertion	Interrater reliability Standardized $r^2 = .88-.98$ 68-91% agreement Personalized $r^2 = .88$ 65-85% agreement	Convergent Behavioral measures during personalized, but not standardized, role-plays generally related to self-report depression scales
Study-Specific Unstructured opposite-sex conversation	Clark & Arkowitz (1975)	High and low social anxious adult men	Global ratings of social skill and social anxiety	Interrater reliability Skill $r^2 = .38$ Anxiety $r^2 = .26$	Discriminative High socially anxious participants were more anxious than low socially anxious participants No difference on ratings of social skill
Study-Specific Unstructured conversation with male	Glasgow & Arkowitz (1975)	High and low frequency dating college students	Behavioral indices of social skill Partner ratings of skill, and anxiety	Interrater reliability $r^2 = .59-.92$	Discriminative No differences between high and low frequency daters on any behavioral measure, or partner ratings of skill and anxiety
Study-Specific Interactions designed to elicit assertion	Gorecki, Dickson, Anderson, & Jones (1981)	High and low assertive college students	Assertiveness and social anxiety	Interrater reliability 91-97% agreement $r^2 = .83-.98$	Discriminative High assertive participants rated as more assertive, more assertive requests, and less anxious than low assertive participants External Assertion in role-play generally greater than assertion in <i>in vivo</i> interaction
Study-Specific Interactions simulating sexual coercion and propositions	Kelly, St. Lawrence, & Brasfield (1991)	High AIDS-risk homosexual adult men	Overall refusal effectiveness and skill responding to sexual propositions	Interrater reliability Refusal assertion $r^2 = .62$ Causal proposition responding $r^2 = .76$	Discriminative Neither refusal assertion nor causal proposition responding contributed to discriminating participants later classified as "safe" or "unsafe" in sexual practices after risk-reduction intervention

Test	Citation	Study population	Social behaviors assessed	Demonstration of reliability	Demonstration of validity
Study-Specific Unstructured group interactions	Lewinsohn, Mischel, Chaplin, & Barton (1980)	Depressed adults, nondepressed psychiatric controls, and nonclinical controls	Behavioral attributes related to social competence	Interrater reliability Pretreatment Spearman Brown $r^2 = .74$ Posttreatment Spearman Brown $r^2 = .68$ Internal Consistency Pretreatment $\alpha = .95$ Posttreatment $\alpha = .97$	Discriminative Depressed participants rated significantly lower on social competence than either psychiatric or nonclinical controls. Depressed participants ratings of social competence improved significantly following treatment, but no change was shown for psychiatric and nonclinical controls.
Study-Specific Structured opposite-sex dating situations	Nelson, Hayes, Felton, & Jarrett (1985)	College students	Global ratings of social skill and social anxiety	Interrater reliability Skills $r^2 = .41-.90$ ($M = .66$) Anxiety $r^2 = .00-.86$ ($M = .16$)	Convergent No difference between participants rated high or low social skill on level of social skill from interview or questionnaire.
Study-Specific Speech to an audience	Newman et al. (1994)	Social phobia: adults with public speaking fears	Global anxiety rating and speech performance rating	Interrater reliability Anxiety $r^2 = .49-.67$ Performance $r^2 = .40-.59$ Test-retest reliability No mean differ- ence in anxiety ratings for control group over 2 weeks, but a significant increase in performance	Discriminative Participants receiving behavioral treatment showed a decrease in rated anxiety, but no change for control group Both control and treatment groups improved on speech quality
Study-Specific Naturalistic or role-played interaction	St. Lawrence, Kirksey, & Moore (1983)	High and low assertive college students	Behavioral measures of assertion	Interrater reliability $r^2 = .49-1.00$	Discriminative High assertive participants rated as more assertive and gave longer responses, but no difference in speech latency, number of refusals, and eye contact External No difference in speech duration or latency, number of refusals, eye contact or overall assertiveness across role-plays and naturalistic interactions

Test	Citation	Study population	Social behaviors assessed	Demonstration of reliability	Demonstration of validity
Study-Specific Opposite-sex interactions	Twentyman & McFall (1975)	"Shy" and "confident" college men	Dating skills and anxiety	Interrater reliability r^2 (pre) = .29-.98 r^2 (post) = .41-.94 Test-retest reliability Skill r^2 = .16-.46 Anxiety r^2 = .00-.31	Discriminative Shy participants were rated as more anxious, less skillful, and engaging in the task for a shorter time than confident participants. Shy participants given behavioral training were less anxious and more skillful than shy controls
Study-Specific Unstructured conversation with female confederate	Walters & Hope (1998)	Social phobia and nonanxious adults	Indices of cooperativeness, dominance, submissiveness, and escape-avoidance	Interrater reliability Mean r^2 = .49-.98 Spearman-Brown r^2 = .73-.99	Discriminative Lower ratings of praise, facing partner, using commands, and bragging for social phobia participants. No other differences. Convergent Facing partner, use of commands, bragging, and interruptions correlated with SPAI-SP.
Study-Specific Same and opposite sex interactions and public speech	Woody, Chambless, & Glass (1997)	Social phobia adults	Combined observer and confederate global ratings of social skill and anxiety	Interrater reliability r^2 = .18-.49 (between observer and confederate ratings) Internal consistency α = .92	Discriminative Appreciable improvement following cognitive-behavioral therapy on ratings of skill and anxiety

Note: r^2 = squared Pearson correlational coefficient; α = Chronbach's alpha; κ = kappa coefficient; G = generalizability coefficient

A few studies have examined the correspondence between analogue and naturalistic observation measured in clinical samples. Curran (1982) noted good correspondence of skill ratings obtained from male psychiatric patients during role-played interactions and unobtrusive naturalistic observations. Bellack, Hersen, and Turner (1979), however, noted moderate relationships between role-played and naturalistic interactions with psychiatric patients on ratings of speech latency, eye contact, and compliance but not on ratings of smiles, praise, number of requests, or overall ratings of assertiveness. Bellack, Hersen, and Lamparski (1979), in a detailed study comparing the behavior of psychiatric patients in role-played heterosocial interactions and staged waiting-room interactions, found only moderate correspondence for women, and minimal correspondence for men. Specifically, only 36% of the relationships between role-played and naturalistic behavior were statistically significant for women, with a strongest correlation of $r^2 = .35$. Only 22% of the relationships between role-played and naturalistic behavior were statistically significant for men, with a maximum correlational coefficient of less than $r^2 = .25$.

Overall, data regarding the external validity of role-play methods are equivocal and certainly not complete. Although there is some evidence for the correspondence between

role-played and naturalistic behavior, there is also a tendency for performance to be somewhat superior in role-played scenarios. This suggests that clinicians should consider that the resulting assessment data may overestimate the client's social functioning. Continued evaluation of role-play methods is necessary to improve the degree of relationship between role-play and naturalistic behavior across different role-play methods and different clinical and nonclinical populations.

Other Aspects of Validity of Analogue Observation Methods

Discriminative Validity

Although evidence for a correspondence between role-played and naturalistic behavior is limited, there is considerable support for the ability to discriminate defined groups using observational data from role-play methods (see Table 1). Comparisons of psychiatric and nonclinical samples consistently show differences between the groups on global ratings of social skill (Curran, 1982; Donahoe et al., 1990) and anxiety (Curran, 1982). Curran (1982) reported that observer rated social skill and anxiety of male schizophrenic inpatients and outpatients differed in the expected direction from observer rated social skill and anxiety of male nonclinical military personnel. Lewinsohn, Mischel, Chaplin, and Barton (1980) reported that observer ratings of social competency obtained during unstructured group interactions effectively discriminated depressed adults from nondepressed adults and psychiatric control participants.

Socially anxious and nonanxious individuals also appear to be discriminable based on analogue observational global ratings of skill (Beidel, Turner, & Dancu, 1985; Merluzzi & Biever, 1987; Twentyman & McFall, 1975; but see Clark & Arkowitz, 1975) and anxiety (Beidel et al., 1985; Clark & Arkowitz, 1975; Twentyman & McFall, 1975). However, when assessing specific behavioral characteristics of anxiety and skill, such as gaze aversion, posture, and speech loudness, differences between socially anxious individuals and nonanxious individuals on behavioral ratings were less consistent (Beidel et al., 1985; Walters & Hope, 1998). Partially supporting the discriminative validity, Alden and Bieling (1998) reported that socially anxious women were rated as less likeable and appropriate than nonclinical women in a negative appraisal situation, but the groups did not differ under positive appraisal conditions. Conflicting evidence of the discriminative validity has been provided by Glasgow and Arkowitz (1975), who found no differences between high and low frequency daters on behavioral ratings of social skill obtained from either observers or the role-play partners.

The utility of role-play methods in discriminating individuals who relapse post-treatment was explored by Kelly, St. Lawrence, and Brasfield (1991). Assertiveness ratings of homosexual men at risk for AIDS who engaged in role-played scenes depicting requests for unsafe sexual practices were obtained following an AIDS prevention program. The assertiveness ratings, however, did not differentiate among individuals who were later classified as sexually safe and those who relapsed to their previous unsafe practices.

In general, the evidence regarding discriminative validity of global ratings of skill and anxiety appears strong. Global ratings effectively discriminate between control samples and individuals with various clinical conditions. However, as noted with comparisons of

socially anxious and nonsocially anxious samples, the discriminative validity of specific behavioral ratings (e.g., gaze, voice qualities, etc.) is less established.

Convergent and Divergent Validity

Several studies have estimated the correspondence between ratings obtained from role-play methods and various other measures (see Table 1). Male psychiatric patients classified as high and low overall on assertiveness, based on performance in the Behavioral Assertiveness Test, differed significantly ($r^2 = .21$) on self-reported assertiveness (Eisler, Miller, & Hersen, 1973). However, they reported no significant difference between the assertiveness groups on self-reported personality. Donahoe et al. (1990) reported that sending skills, but not receiving or processing skills, were significantly related to IQ among participants with schizophrenia ($r^2 = .13-.20$), whereas Curran (1982) reported that Simulated Social Interaction Test skill ratings were positively related to skill ratings obtained from a clinical interview with male psychiatric patients.

Convergent validity has also been examined using analogue assessment methods with samples of socially anxious individuals. Behavioral ratings of social performance obtained during an unstructured conversation between individuals with social phobia or nonanxious participants and a female conversation partner were generally related ($r^2 = .01-.24$) to self-reported social anxiety (Walters & Hope, 1998). Merluzzi and Biever (1987) assessed ratings of social anxiety and skill obtained during structured and unstructured opposite-sex interaction with a sample of male college students classified as high anxiety, low skill; medium anxiety and skill; and low anxiety, high skill. They noted moderate relationships between independent observer global ratings of social anxiety and skill and ratings of anxiety and skill obtained from the role-play partners ($r^2 = .06-.23$).

Similarly, Kern (1991) reported that overall assertiveness ratings obtained from college students engaging in the Ideographic Role-Play test were positively correlated ($r^2 = .17-.18$) with self-reported assertiveness. Supporting the divergent validity of the Ideographic Role-Play test, overall ratings of assertiveness showed no significant relationship to a measure of social desirability (Kern, 1991). Contrary to the preceding findings, Nelson, Hayes, Felton, and Jarrett (1985) found no evidence for discriminative validity during a role-played structured dating situation with college students. They reported that college students classified as high and low in assertiveness based on observer ratings during the role-play did not differ on assertiveness ratings obtained from either clinical interview or self-report questionnaire.

Finally, partial evidence for the convergent validity of analogue assessment methods employed with depressed individuals has been reported by Chiauzzi, Heimberg, Becker, and Gansler (1985). They found that self-reported depression was generally related to behavioral indices of assertion during individualized role-plays. However, Chiauzzi et al. (1985) reported no significant relationship between self-reported depression and behavioral indices of assertion obtained from standardized role-plays.

Some, but not all, of the reported studies suggested that data from analogue observational methods tend to relate to measures of similar constructs. However, apart from the Kern (1991) study, no evidence concerning divergent validity was reported. Clearly, more research is required to establish the degree of correspondence between ratings derived

from analogue observational methods and related measures of the constructs of interest, as well as divergence from measures of unrelated constructs.

Treatment Sensitivity

A dimension particularly important in the clinical utility of any assessment method is the sensitivity of the method to detect clinically significant changes in target behavior(s). The treatment sensitivity of analogue assessment methods has been noted in several studies of individuals receiving treatment for social anxiety disorder (see Cohn & Hope, 2000, for a review). For example, Woody, Chambless, and Glass (1997) noted "appreciable improvement" on observer ratings of anxiety and skill during opposite and same-sex interactions and a public speech following cognitive-behavioral treatment with a sample of adults with social phobia. Newman, Hofmann, Trabert, Roth, and Taylor (1994) also noted improvement on observer ratings of anxiety during a public speech following treatment for social anxiety disorder. Participants not receiving treatment showed no significant change in anxiety ratings. Similar treatment effects were obtained from participant ratings of improvement and several self-report measures of social anxiety. In comparing the effectiveness of cognitive-behavioral group treatment and behavioral group treatment for social anxiety disorder to a no-treatment wait-list control group, Hope, Heimberg, and Bruch (1995) found that observer and self-ratings of performance during an individualized anxiety-provoking role-play increased significantly following receipt of either treatment. No significant change was noted for a wait-list control group.

Much like reports of individuals receiving treatment for social anxiety, there is evidence that observer ratings are sensitive to assertiveness training effects. Baggs and Spence (1990) found that women receiving an eight-week assertiveness training program improved significantly on observer ratings of speech latency ($r^2 = .12$), compliance ($r^2 = .27$), and overall assertiveness ($r^2 = .40$) during a structured interaction, whereas no significant change was apparent for the control participants. Furthermore, evidence of improved assertiveness following treatment was obtained using self-report questionnaires and diary reports, suggesting a correspondence between assessment methods. Glueckauf and Quittner (1992) found that physically disabled adults receiving assertiveness training showed improvements from pre- to posttreatment in observer ratings of directed looking ($r^2 = .23$), assertive responding ($r^2 = .30$), and appropriate affect ($r^2 = .13$). Physically disabled adults in a control group did not change on any observer ratings from pre- to posttreatment. These effects mirrored changes in self-reported assertiveness, with the treatment group reporting improvements, and no change being reported by control participants. Interestingly, the treatment sensitivity was questionable for the maintenance of treatment gains, as ratings from role-played interactions showed a return to pretest levels whereas self-reported assertiveness suggested maintenance of treatment gains.

Finally, treatment sensitivity has been noted with samples of depressed adults. Following a depression treatment program, the social competence ratings of depressed adults showed significant improvement, whereas nondepressed control participants showed no change over time (Lewinsohn et al., 1980). This evidence was further established in that Minnesota Multiphasic Personality Inventory (MMPI) *D* scale score decreases confirmed the effectiveness of the depression treatment.

The sensitivity of role-play methods to treatment effects appears to be quite well established. The available data indicate that behavioral ratings improve over the course of treatment for social anxiety, unassertiveness, and depression, with posttreatment individuals being discriminable from individuals who have not received treatment. Furthermore, several studies (Baggs & Spence, 1990; Glueckauf & Quittner, 1992; Hope, Heimberg, & Bruch, 1995; Lewinsohn et al., 1980; Newman et al., 1994) suggest that improvements in behavioral ratings tend to mirror improvement data obtained from self-report data sources.

Estimates of Reliability for Analogue Observational Methods

Interrater Reliability

Perhaps the most frequently reported psychometric dimensions in studies employing role-play methodology are estimates of interrater reliability. As shown in Table 1, the majority of these studies estimate the interrater reliability to range from good to excellent. Although most studies use highly trained raters (e.g., Bellack, Hersen, & Turner, 1979), Hope and colleagues (Hope & Heimberg, 1988; Hope, Heimberg, & Bruch, 1995) have reported excellent interrater agreement with untrained raters using global ratings. When molecular ratings employ anchored rating scales (e.g., Caballo & Buela, 1988), interrater reliability is augmented (Bellack, 1983).

Test-Retest Reliability

Only one study (Donahoe et al., 1990) directly assessed the test-retest reliability of ratings obtained from role-plays, reporting a very high temporal stability with a sample of schizophrenic outpatients ($r^2 = .77$ to 1.00). Twentyman and McFall (1975) reported that the pre-treatment and posttreatment skill ratings of "shy" men not receiving behavioral treatment were moderately correlated ($r^2 = .16$ – $.46$), but anxiety ratings were essentially unrelated ($r^2 = .00$ – $.31$). Thus, the test-retest reliability of observer ratings from role-play methods is somewhat unclear, as there is evidence of variability in both anxiety and skill ratings across time. Given the variability in human behavior, and the possibility that the symptomatology of clinical patients may wax and wane, some temporal incongruence must be expected. Therefore, clinicians employing role-play methods to assess social functioning should consider using multiple assessments over time to strengthen confidence in role-play data.

Methodological Considerations in Using Analogue Assessments of Social Behavior

As was evident in the preceding discussion of the validity of analogue observational assessment, changes in the methodology impact clients' reactions to the role-play situation. Also, different methods for recording and coding data may have implications for interpretation of the data. The various issues to consider when conducting these assessments will be divided into two general areas for the purpose of discussion. First, *procedural components* that may impact on the observed individual are discussed. Such components include instructions given to the participant, the use of confederates and their personal characteristics, and personalization or familiarity of role-played scenes. Second, *structural components* not directly impacting the participant during the observation are discussed. Examples of

structural components are molar versus molecular-level ratings and the use of recording technologies or live observers.

Procedural Components

Instructions

The nature of the instructions provided to clients prior to a role-play task may considerably influence the quality of performance exhibited during the assessment. Several instruction variations have been employed in role-play assessment, including “act as you normally do/would” and “act as you believe the most skilled person would act.” Others (e.g., Kern, 1982) have instructed clients to replicate previously videotaped waiting room interactions involving the client. Individuals adjust their behavior based on the instructions, with better performance being associated with high demand instructions to “act as you believe a very assertive person would act” versus low demand instructions to “act as you normally do” (Nelson et al., 1985; Nietzel & Bernstein, 1976). However, high demand instructions may not bring individuals with poor social skills into the normal range as Nelson and colleagues reported that low assertiveness participants did not achieve the same level of performance as demonstrated by highly assertive participants under either high- or low-demand conditions.

Clinicians employing role-play assessment methods should give consideration to the objectives of the assessment and script the instructions accordingly. Assessments in which the focus is solely an evaluation of a client’s current skill performance or anxiety should employ low-demand instructions. Assessments conducted to test the possible limits of an individual’s social skills repertoire could employ a sequential combination of high- and low-demand instructions to identify discrepancies between performance and capability.

Individualized and standardized role-play scenarios

Both standardized and individualized role-plays have been used in analogue observational assessment, with advantages and disadvantages for each, depending upon the goal of the assessment. Standardized role-plays facilitate comparisons to the performance of other people. For example, a clinician regularly employing a standardized set of role-play scenarios for individuals with social anxiety may develop some local norms to gauge the degree of anxiety and impairment relative to other socially anxious clients. Use of published role-play assessments with standardized scenarios allows comparison to normative data for certain clinical and nonclinical populations. Despite these advantages, the scenarios in standardized role-plays may not be the best fit for a particular individual, resulting in insufficient or misleading data. Individualized role-play scenarios, on the other hand, may have higher convergent validity (Chiauzzi et al., 1985; Torgrud & Holborn, 1992) and can more closely match a particular client’s presenting concerns. Inclusion of important personal and cultural aspects such as sexual orientation, gender roles, social mores, and socioeconomic status ensures the assessment is relevant to a particular client’s life circumstances. Greater accuracy is achieved if a client’s performance is not artificially impaired due to unfamiliar or awkward scenarios. These relative advantages and disadvantages for

standardized and individualized scenarios suggest that seeking convergence from a combination of scenarios may be the best strategy.

Stimulus format

Perhaps one of the most varied and frequently debated aspects of using analogue methods for assessment of social behavior is the stimulus format. Early studies tended to use single prompts or statements, but the trend is toward multiple interrelated prompts that may or may not be standardized. Few social interactions comprise a single prompt and a single response. Rather, social interactions tend to involve a series of interrelated prompts, responses, and counter-responses that allow the interactors to fully convey their messages. Indeed, as early as 1983, Bellack suggested that the use of the "single prompt role play yields extremely limited and stilted responses . . . and is no longer justified" (p. 38). Furthermore, anecdotal (Bellack, 1983) and empirical (Kirchner, Kennedy, & Draguns, 1979) evidence suggested that differences in social behavior might not become apparent until after several exchanges.

The extent to which the prompts can or should be standardized must also be considered when using analogue observational methods. Standardized prompts protect the internal validity of the assessment procedure by maintaining ongoing stimulus equivalence across participants. However, external validity may be compromised, as standardized prompts may not realistically capture a logical sequence of communication between two parties. Nonstandardized prompts limit comparisons between individuals or across observation occasions as different follow-up prompts may generate responses differing in content, length, or provoked anxiety. Many clinical researchers appear to have aimed for a compromise by standardizing the style of the interaction partner (e.g., affect, friendliness, timing rules for breaking silences) but allowing a naturally flowing interaction (e.g., Edelman & Chambless, 1995). As noted earlier, this method appears valid with various diagnostic groups.

Use of confederates and role-play partners

By definition, social behavior involves contingent or noncontingent interaction that requires one or more role-play partners for any analogue observational assessment. In research studies the role-play partners are often trained research assistants (e.g., Alden & Bieling, 1998; Bellack, Hersen, & Turner, 1979) or other participants (e.g., Bellack, Hersen, & Lamparski, 1979; Kelly et al., 1991). In some larger settings, hospital or agency staff and colleagues may serve this function. However, such assistants may be unavailable in some settings or add considerable cost and inconvenience to the assessment. Although the clinician conducting the assessment may serve as the role-player (e.g., Hope & Heimberg, 1993), there is no research on whether this impacts the outcome of the assessment or any subsequent therapeutic alliance. Certainly the unique nature and power differential of the client-clinician relationship may influence the client's behavior in the role-play. Anecdotal evidence suggests clinicians commonly serve as role-play partners in clinical practice (e.g., Hope, Heimberg, Juster, & Turk, 2000), but clinical judgment should be used to determine when this is appropriate.

Role-play partner characteristics

Personal characteristics of role-play partners may influence a client's behavior during the assessment (Torgrud & Holborn, 1992). Although there is little direct research on this issue, several researchers have acknowledged its importance by offering brief physical descriptions of confederates used in a study (e.g., Merluzzi & Biever, 1987). Eisler, Hersen, Miller, and Blanchard (1975) noted various differences in speech latency and dysfluencies when male psychiatric patients interacted with male versus female confederates in role-plays requiring positive assertion. This suggests multiple role-plays with different confederates may be ideal, as behavior may not generalize across role-play partners. At the very least, clinicians should consider the personal characteristics of the role-play partner when interpreting the data.

Structural Components*Molar-level assessment*

Molar assessments that involve making a global evaluation of the quality of the social behavior performed by the observed participant have been commonly used in published analogue observational assessment. As noted earlier, interrater reliability of molar-level ratings appears to be relatively consistently high regardless of the degree of training received by raters. Despite the fact that molar assessments are assumed to be global ratings of social skills and functioning, evidence suggests that molar ratings of social skill tend to predominantly be based upon verbal content and gaze (Bellack, 1983; Conger & Farrell, 1981; Galassi et al., 1976; Romano & Bellack, 1980; Trower, 1980). Thus, global ratings may not accurately capture the performance of someone with significant deficits in other behaviors such as nonverbal gestures, verbal volume, and intonation.

Although molar ratings of social behavior can be used as a gross screening procedure or to monitor overall progress with an intervention, their utility in treatment planning is limited. Interventions such as social skills training require data regarding specific performance deficits such as duration of gaze, verbal response length, or severity of visible anxiety symptoms. For these data, observations using molecular assessment will need to be employed.

Molecular-level assessment

Molecular rating of social behavior involves the observation and recording of specific social behaviors or anxiety symptoms. Examples of typical behaviors to rate include eye contact or gaze; response latency and duration; speech content, inflections, volume, and dysfluencies; and bodily movements such as tremors, self-manipulations, gestures, and smiles. Frequently, molecular ratings assess the frequency, duration, or intensity of the specific behaviors, with the assumption that larger or smaller amounts (of frequency, duration, intensity, etc.) indicate better or worse social functioning. However, optimal social behavior sometimes follows an "inverted U" pattern with a moderate frequency or intensity being the most skilled (i.e., both no eye contact and staring are inappropriate; Bellack, 1983; Trower et al., 1978). Unfortunately, many of the standardized role-play assessments available do not account for molecular behavior that occurs at both extremes.

The relationship between molecular and molar ratings has been controversial. According to Bellack (1983), specific social behaviors are “presumed to be the basic elements of interpersonal communication, which together comprise the social skill construct” (p. 33). Conversely, however, Conger and Conger (1982) suggested that social behavior is “greater than the sum of its parts” (p. 328). Despite advances in our understanding of social behavior, this issue has not been resolved. The collection of both molar and molecular ratings may be the best solution. (See Meier & Hope, 1998, for a full discussion.)

Client attractiveness

In his 1988 review of the relationship between social functioning and physical attractiveness, Calvert concluded that physical attractiveness may confound ratings of social skill such that physically attractive people are rated as more skilled than less physically attractive people. Furthermore, improvements in social skill resulted in higher ratings of physical attractiveness among developmentally disabled adults (Mueser, Valenti-Hein, & Yarnold, 1987). However, Merluzzi and Biever (1987) and Wessberg et al. (1979) both reported that global anxiety and skill ratings of college students were not related to observer ratings of participant attractiveness. In the one noncorrelational study to examine the relationship between social skill and attractiveness, Hope and Mindell (1994) found that attractive individuals may be seen as more socially skilled, but only when skill performance meets a prerequisite level of competency. An attractive but unskilled individual was not perceived as more skilled than a less attractive, unskilled individual.

Immediate versus delayed (recorded) assessment

Video-recording is commonly employed in observational research of social behavior. Although video equipment is less common in nonresearch settings, recording analogue assessments has many advantages including the opportunity for the clinician to review the tape to make various types of ratings, the possibility of ratings from objective observers, and the possibility of using the tape for feedback to the client. As video cameras become more ubiquitous in our daily lives, recording may become less anxiety provoking for the client. Improved technology and decreasing costs may improve feasibility.

Summary

The preceding sections highlighted important procedural and structural considerations in developing and employing analogue methods for the assessment of social functioning. Variations in each of these factors could have a dramatic impact on the validity and reliability of the obtained assessment data. Notably, variations in instructional demand can greatly influence subsequent behavioral performance, thereby impacting the external validity of the data. As well, individualization of role-played scenarios and prompts potentially augments the external validity of the role-plays by tailoring them to the client's natural environment but limits the extent to which conclusions can be based upon local or published norms. Conversely, standardization of role-played scenarios and prompts allows for such norm-based comparisons but may limit the external validity of the data as the scenarios and prompts may not be typical or realistic for any given client. These and

the other considerations discussed earlier, such as the use of single prompts or multiple prompts, the use of role-play partners, immediate versus delayed or recorded assessment, and the personal characteristics of both the role-play partners and the clients themselves, can individually and mutually bear on the validity of inferences derived from analogue methods for assessing social functioning. Similarly, the level of assessment data (molar or molecular) collected using analogue methods allows for different clinical conclusions, each of which may be differentially valid and useful for various clinical populations and assessment questions. Molar-level assessments provide a global assessment of social functioning that may serve as an index of treatment outcome, whereas molecular-level assessments are particularly useful in identifying strengths and deficits that may become specific treatment targets. Furthermore, all of these factors, particularly the level of assessment data collected, may bear on the reliability of the assessment data collected. Therefore, it is of critical importance to the clinician to carefully consider these procedural and structural factors to ensure that the assessment provides valid and reliable responses to the specific assessment questions for a given client.

Standardized Analogue Observation Methods for Social Behavior Assessment

As noted earlier, several standardized analogue observational methods have been developed for the assessment of social behavior. By *standardized* methods, we are referring to published assessment methods with demonstrated scoring procedures, established scenarios or techniques for devising individualized scenarios, and detailed procedural instructions. These methods, and estimates of their reliability and validity, are summarized in Table 1. For clinicians, these standardized assessments have the advantage that they are prepared, relatively easy to arrange and use, and tend to have demonstrated reliability and validity. Despite this, the fact that they are structured and prepared may reduce the degree to which they are appropriate to the specific situational and personal characteristics of the client.

Behavioral Assertiveness Test-Revised (BAT-R)

In the Behavioral Assertiveness Test-Revised (BAT-R; Eisler et al., 1975), male clients are presented with 32 standardized situations that are role-played with male and female confederates. Half of the situations involve interacting with a familiar individual (e.g., boss, spouse), while half involve interacting with an unfamiliar individual (e.g., waiter/waitress). Additionally, half of the situations require positive assertion skills, such as praising another, while the remainder require negative assertion skills, including refusal or voicing displeasure.

A client is instructed to "act as he typically does." The scene is narrated, followed by the role-play partner delivering a predetermined prompt to the client. All of the role-plays are video-recorded for later scoring. Scoring incorporates a molar rating of overall assertiveness as well as several molecular ratings of eye contact, frequency of smiles, and duration of eye contact. Finally, specific behaviors related to negative (e.g., compliance, requests for new behavior) and positive (e.g., expressing praise, expressing appreciation) social behavior are rated on an occurrence or nonoccurrence basis.

As seen in Table 1, interrater reliability tends to be good when employing the BAT-R with psychiatric individuals (Eisler, Hersen, Miller, & Blanchard, 1975; Bellack, Hersen, & Turner, 1979). However, interrater reliability may be weaker when assessing unassertive, nonclinical individuals (Baggs & Spence, 1990). Evidence for the external validity of the BAT-R is questionable, as Bellack, Hersen, and Turner (1979) noted only moderate to weak correspondence between BAT-R assessed behaviors and similar behaviors in staged naturalistic interactions. However, BAT-R molar ratings of overall assertiveness appear to converge with several molecular behavior ratings (Eisler, Hersen, Miller, & Blanchard, 1975) and assertiveness ratings obtained from structured interviews (Bellack, Hersen, & Turner, 1979). Finally, there is evidence that BAT-R overall assertiveness ratings, response latency, and compliance ratings are sensitive to treatment effects (Baggs & Spence, 1990).

One advantage of the BAT-R is that it assesses a range of heterosocial interactions including interactions with familiar and unfamiliar individuals and interactions requiring positive and negative assertion. However, given that it is only designed for use with men, and only recreates heterosocial interactions, the BAT-R appears to be limited in its clinical utility. Furthermore, the interrater reliability of the BAT-R in assessing nonclinical populations remains questionable and awaits further investigation.

Assessment of Interpersonal Problem-Solving Skills (AIPSS)

The Assessment of Interpersonal Problem-Solving Skills (AIPSS; Donahoe et al., 1990), a tool designed to assess social skill performance and social competence, employs 14 videotaped interactions that are individually presented to the client. The first scene orients the client to the task, 10 scenes portray an interpersonal problem, and 3 scenes have no identifiable interpersonal problems. Clients are instructed to identify with a specific actor in the video segment. After viewing each scene, clients are asked to discriminate if there is a social problem in the scene. If a problem is identified, the client is asked to describe the problem and report what he or she would do in the scene to rectify the problem. Finally, the client role-plays his or her proposed solution with the experimenter or a trained assistant. The entire AIPSS is video-recorded and later scored by highly trained observers using a structured scoring manual (Donahoe, Carter, Bloem, & Leff, 1984).

The AIPSS utilizes a complex scoring procedure to assess the participant's ability to identify an interpersonal problem, to develop and describe a solution to the problem, and to enact the solution (Donahoe et al., 1990). The participant's performance on the AIPSS is assessed along six scales, clustered into three domains: Receiving Skills (identification and description), Processing Skills (processing), and Sending Skills (content, performance, and overall role-play performance). Furthermore, two scoring systems are used that account for whether problems were accurately identified.

Estimates of the reliability of the AIPSS appear adequate (see Table 1). Donahoe et al. (1990) reported acceptable interrater and test-retest reliability in assessments of schizophrenic and nonclinical men when using either the general or specific scoring procedures. The internal consistency of the AIPSS subscales was variable, ranging from $\alpha = .57$ to $\alpha = .93$, suggesting that refinement of some of the items may be warranted. There is support for the discriminative validity of the AIPSS, as ratings from the general-scoring system effectively differentiated nonclinical participants and participants with schizophrenia.

Despite the utility of the AIPSS, administration, and scoring is complex, requiring considerable investment in training for role-play partners and observers. Consequently, the AIPSS would appear to be most appropriate for a large-scale institution, which regularly assesses and treats a client population characterized by social difficulties. Conversely, the AIPSS does not appear to be the most cost-effective and accessible tool for private clinics or smaller outpatient settings.

Simulated Social Interaction Test (SSIT)

The Simulated Social Interaction Test (SSIT; Curran, 1982) assesses clients' social skill and anxiety using 12 structured role-played scenarios with a single prompt format. Following 4 practice scenarios, the 8 scored scenarios assess situations involving disapproval/criticism, social visibility/assertiveness, confrontation/anger expression, heterosocial contact, intimacy/interpersonal warmth, conflict/rejection by parents, interpersonal loss, and positive emotional expression. In each role-play, a narrator verbally outlines the scenario (e.g., "You have had an argument with a close friend. She says to you . . ."), at which point the role-play partner delivers a single predetermined prompt (e.g., "I don't want to talk about it anymore. I'm leaving."). The client then delivers a response and the scenario is terminated. Although the single prompt design was selected to maximize standardization and make the instrument easy to use, (Curran, 1982), single prompts may not be adequate, as discussed earlier. As outlined in Curran (1982), each judge independently makes separate ratings of the client's social skill and anxiety in each situation on anchored 11-point Likert-type scales.

Curran et al. (1980) reported very high generalizability coefficients in support of the reliability of the SSIT. Furthermore, as reported in Table 1, SSIT ratings were moderately correlated with ratings obtained from naturalistic observations of male psychiatric patients, supporting the external validity of the SSIT. Curran (1982) also provided evidence of the convergent and discriminative validity of the SSIT when employed with a sample of psychiatric patients and nonclinical military personnel.

Overall, the SSIT appears to be a valid and useful method for the observational assessment of social skill and anxiety among adult male psychiatric patients. As noted earlier, the SSIT was designed for maximal efficiency and ease of use in clinical settings. However, clinicians wishing to employ the SSIT should be cautioned that these design advantages might be offset by the use of single-prompt methodology.

Social Skill Behavioral Assessment System (SSBAS)

Caballo and Buela (1988) constructed the Social Skill Behavioral Assessment System (SSBAS) in an effort to develop a valid role-play test of social skill that incorporates both molar ratings of functioning and assessment of an empirically identified set of 21 molecular behaviors. Only one role-play is performed—a 5-min heterosocial interaction with a trained confederate in an unstructured casual conversation. Judges rate eight molar ratings of global functioning along a 7-point scale, while molecular behaviors are rated on 5-point scales of the behavior's adequacy. The molecular ratings, however, assume linearity from deficient behavior to adequate performance without accounting for behavioral excesses (see Bellack, 1983).

Clients are brought into a furnished room and introduced to an unfamiliar confederate of the opposite sex. The client and confederate are instructed to interact and to get to know each other over a 5-min period that will be videotaped. Confederates are trained to adhere to specific requirements, such as waiting 20 sec before initiating a conversation, and are told the frequency with which to look at the client or engage in socially reinforcing behaviors such as nodding or smiling.

The authors reported obtaining very high estimates of interrater reliability when assessing college students with the SSBAS (Caballo & Buela, 1988; see Table 1). They also reported that the molar ratings of social skill provided by both the independent observers and the participants were, in general, moderately correlated with molecular behavior ratings obtained from the SSBAS. However, observer and participant global ratings of social anxiety were generally poorly correlated with molecular ratings.

The SSBAS appears to be a very useful and comprehensive role-play method. The combination of an overall rating of social skill and molecular behaviors offers the clinician the ability to assess the presence of impaired social functioning as well as detect specific behaviors that may be underlying or mediating the social difficulties. Furthermore, the brief nature of the SSBAS—one 5-min interaction—makes it appealing for use in a standard clinical session. However, in using only one brief interaction, the sensitivity of the SSBAS in detecting difficulties in social functioning may be compromised. For example, the casual unstructured interaction may not generate useful data regarding individuals who have difficulty interacting with persons in positions of authority.

Ideographic Role-Play Test (IRP)

The IRP (Kern, 1991) represents a compromise between standardized role-plays wherein interactions may or may not represent situations experienced by participants, and individualized role-plays designed specifically for use with certain clients. Rather than develop interactions that the participants must perform, Kern developed six general assertion situation types. Participants are asked to recall six examples of each situation type that they have recently experienced. Considerable detail is obtained regarding the specifics of the situations, the relationship between the participant and the other individual involved in the interaction, and characteristics and behaviors of the other person. Example situations deemed to be appropriate are then role-played with a partner using any props available. Role-played interactions are kept brief (2 to 6 exchanges). Situation types involve (a) not wanting to lend an item that someone has asked to borrow, (b) buying something that turns out not to be what was wanted, (c) being requested to do something undesirable, (d) receiving a solicitation to purchase an unwanted item, (e) someone doing something that disturbs the participant, and (f) wanting another person to do something he or she promised to do previously.

Following development of each acceptable interaction, the interaction is role-played with an experimenter or assistant. An observer rates the participant's overall assertiveness in each interaction on a 6-point scale ranging from total assertion (1) to total submissiveness or aggressiveness (6). Scores are summed across all valid role-played interactions within each situation type, yielding scores for six broad areas of assertiveness. Further, all scores are summed into a total rating of assertiveness. For both situation type scores and

the overall score, if fewer than six interactions are role-played for any situation type (i.e., participant cannot recall six recent examples of undesirable requests), average situation type scores are substituted for missing data.

Kern (1991) reported excellent interrater reliability estimates for individual situation ratings and the overall assertiveness rating, although situation-type scores showed considerable variability in interrater reliability. Internal consistency, however, was low when assessed among situations and overall. IRP situation scores were moderately correlated with behavioral indices of assertiveness obtained during a contrived telephone conversation, providing evidence of the external validity of the IRP. Kern (1991) also provided strong support for the convergent and divergent validity of the IRP, as assertiveness ratings were moderately correlated with self-reported assertiveness but unrelated to measures of social desirability.

The IRP appears to be a novel role-play approach, balancing the need for consistency and standardization with attention to the individual history and social environment of the client. Further, Kern (1991) reported that the IRP holds encouraging psychometric properties. However, because clients must generate example interactions from their recent history, the IRP may not be an ideal role-play method for use with some clients. For example, clients experiencing distortions in reality or those who are highly socially avoidant may not be able to generate sufficient adequate interactions.

Recommendations for Clinical Application

Although the correspondence between analogue and naturalistic observational assessment of social behavior is not as high as one might prefer for clinical decision making, the research literature is quite limited for such a complex question. Rather than asking whether analogue observational assessment has high validity, it would be more fruitful to examine under what conditions (e.g., instructional sets, clinical populations, role-play situations) analogue assessment corresponds to naturalistic observation. As noted earlier, analogue observational assessment generally can distinguish between disordered or nondisordered and treated or untreated individuals. Less is known about convergence with other measures. Therefore, the clinician is cautioned to bear in mind the various procedural and structural influences that can impact the assessment data. With regard to the standardized published role-play assessments, several recommendations for use can be tentatively made.

In psychiatric hospitals and long-term care facilities serving individuals with severe and persistent mental illness, the AIPSS would be most highly recommended for use. The AIPSS provides a wealth of individual data regarding the client's receiving, processing, and sending abilities that can inform treatment planning, and it has demonstrated discriminative validity with schizophrenic samples. However, the AIPSS is a highly complex assessment tool and requires considerable preparation and training. As such, it would only be recommended for institutions with the necessary staff and resources. Alternatively, the SSIT is designed to be clinically efficient and simple and has been validated with psychiatric inpatient populations.

The SSIT appears to be useful in both inpatient hospital settings and outpatient or day-treatment settings for chronically mentally ill clients. As noted earlier, the SSIT is designed

for maximal ease of use, and it has demonstrated strong psychometric properties. However, several aspects that may be of value to clinicians, including rating of molecular-level behaviors and extended interactions with multiple prompts, have been omitted to make the SSIT more clinically accessible.

As well, the BAT-R and IRP would be recommended for outpatient and day-treatment centers for chronically mentally ill clients. The BAT-R has excellent reliability and validity for use with male psychiatric patients. However, without modification (see Baggs & Spence, 1990), it would be of little value for use with women. Because the BAT-R focuses on assertive behavior, it is only appropriate when an assessment of assertiveness is desired, as opposed to a broader assessment of social functioning. The IRP appears to be a promising tool that provides a balance between standardization and ideographic assessment. However, as noted earlier, the IRP may not be useful with more severely impaired or socially isolated populations who may lack the repertoire of recent social experiences necessary to effectively employ the IRP.

Finally, counseling programs and private practices serving nonpsychotic clients would be advised to consider the use of the IRP, SSBAS, BAT-R, and possibly the SSIT. The SSBAS emphasizes skill and anxiety in heterosocial interactions and was found to be reliable and valid for use with a university student population. However, as with the BAT-R, the SSBAS would be less beneficial should a broad focus of assessment be required.

Conclusions

Although it is easy to conclude that more research is needed to further our understanding of the analogue assessment of social behavior, our current knowledge offers guidance to clinicians. Dysfunctional social behavior is a key aspect of many mental health problems. Pretreatment documentation of those problems can be both time and cost efficient. With analogue observational methods, the behavioral data can stand alone or, preferably, be combined with self-report or more traditional assessment strategies to seek converging evidence in three arenas: (a) identification and documentation of the nature and severity of the presenting problem, (b) monitoring progress through repeated assessment, and (c) documenting the effectiveness of interventions. Thus, analogue observational assessment can serve the interests of the various constituencies in the twenty-first century therapeutic enterprise: clients, clinicians, agencies, and third-party payers.

The empirical data are quite supportive of the discriminative validity, convergent validity, and treatment sensitivity of role-play methods in assessing social skill and anxiety. Despite this, the evidence concerning the external validity of role-play methods is equivocal. Thus, the basic question, "Does the behavior exhibited in role-plays correspond to behavior observed in naturalistic situations?" cannot yet be conclusively answered. Given that the use of role-play methods is so pervasive in behavioral assessment and treatment, it is imperative that continued research be conducted to further address this question.

One definitive conclusion that can be drawn from the empirical literature is that the development of good methods for the assessment of social skill and anxiety is not a casual undertaking. Variations in factors such as instructions, number of prompts, selection of

confederates and their characteristics, and the individualization of the role-played scenarios have all been shown to impact the resulting assessment data. Indeed, a simple encouraging comment such as “just do your best” could increase the instructional demand and impair the external validity of the data. Thus, clinicians employing role-play methods must clearly consider each of these factors to develop role-plays that adequately match the individual client and the relevant assessment questions. Similarly, the level of analysis (molar vs. molecular data) and the specifically observed behaviors must be carefully considered in relation to the assessment questions and the purpose of the assessment. Molar-level data tend to be more reliable and can be assessed relatively quickly, whereas molecular-level data provide clinically useful data that can be used to identify behaviors requiring specific therapeutic attention. However, to effectively assess the plethora of behaviors involved in social functioning, recording equipment would certainly be required.

Given the investment clinicians must make to develop sound role-play methods, it is not surprising that standardized methods may be most attractive to practicing clinicians. Indeed, the fact that the aforementioned structural and procedural factors have already been considered by these methods’ authors makes these tools user friendly. Despite this, careful attention must be paid in selecting a standardized tool that is psychometrically sound for the specific client, the assessment questions, and the resources and constraints of the clinical setting.

Future research should address the conditions under which ideographic and standardized role-plays have strong psychometric qualities when used in nonresearch settings. In particular, close attention should be paid to establishing procedures that generalize to clients’ everyday lives and are sensitive to treatment gains.

References

- Alden, L. E., & Bieling, P. (1998). Interpersonal consequences of the pursuit of safety. *Behaviour Research and Therapy, 36*, 53–64.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Andrasik, F., Heimberg, R. G., Edlund, S. R., & Blankenberg, R. (1981). Assessing the readability levels of self-report assertion inventories. *Journal of Consulting and Clinical Psychology, 49*, 142–144.
- Baggs, K., & Spence, S. H. (1990). Effectiveness of booster sessions in the maintenance and enhancement of treatment gains following assertion training. *Journal of Consulting and Clinical Psychology, 58*, 845–854.
- Beidel, D. C., Turner, S. M., & Dancu, C. V. (1985). Physiological, cognitive and behavioral aspects of social anxiety. *Behaviour Research and Therapy, 23*, 109–117.
- Bellack, A. S. (1983). Recurrent problems in the behavioral assessment of social skill. *Behaviour Research and Therapy, 21*, 29–41.
- Bellack, A. S., Hersen, M., & Lamparski, D. (1979). Role-play tests for assessing social skills: Are they valid? Are they useful? *Journal of Consulting and Clinical Psychology, 47*, 335–342.
- Bellack, A. S., Hersen, M., & Turner, S. M. (1979). The relationship of role playing and knowledge of appropriate behavior to assertion in the natural environment. *Journal of Consulting and Clinical Psychology, 47*, 670–678.

- Caballo, V. E., & Buela, G. (1988). Molar/molecular assessment in an analogue situation: Relationships among several measures and validation of a behavioral assessment instrument. *Perceptual and Motor Skills*, *67*, 591–602.
- Calvert, J. D. (1988). Physical attractiveness: A review and reevaluation of its role in social skill research. *Behavioral Assessment*, *10*, 29–42.
- Chiauzzi, E. J., Heimberg, R. G., Becker, R. E., & Gansler, D. (1985). Personalized versus standard role plays in the assessment of depressed patients' social skill. *Journal of Psychopathology and Behavioral Assessment*, *7*, 121–133.
- Clark, J. V., & Arkowitz, H. (1975). Social anxiety and self-evaluation of interpersonal performance. *Psychological Reports*, *36*, 211–221.
- Cohn, L., & Hope, D. A. (2000). Treatment of social phobia: A treatments by dimensions review. In S. Hofmann & P. DiBartolo (Eds.), *From social anxiety to social phobia: Multiple perspectives* (pp. 354–379). Needham Heights, MA: Allyn & Bacon.
- Conger, J. C., & Conger, A. J. (1982). Components of heterosocial competence. In J. P. Curran & P. M. Monti (Eds.), *Social skills training: A practical handbook for assessment and treatment* (pp. 313–347). New York: Guilford Press.
- Conger, J. C., & Farrell, A. D. (1981). Behavioral components of heterosocial skills. *Behavior Therapy*, *12*, 41–55.
- Curran, J. P. (1982). A procedure for the assessment of social skills: The Simulated Social Interactions Test. In J. P. Curran & P. M. Monti (Eds.), *Social skills training: A practical handbook for assessment and treatment* (pp. 313–347). New York: Guilford Press.
- Curran, J. P., Monti, P. M., Corriveau, D. P., Hay, L. R., Hagerman, S., Zwick, W. R., & Farrell, A. D. (1980). The generalizability of a procedure for assessing social skills and social anxiety in a psychiatric population. *Behavioral Assessment*, *2*, 389–401.
- Donahoe, C. P., Carter, M. J., Bloem, W. D., Hirsch, G. L., Laasi, N., & Wallace, C. J. (1990). Assessment of interpersonal problem-solving skills. *Psychiatry*, *53*, 329–339.
- Donahoe, C. P., Carter, M. J., Bloem, W. D., & Leff, G. (1984). *Assessment of interpersonal problem-solving skills*. Unpublished test manual, West Los Angeles Veterans Affairs Medical Center.
- Edelman, R. E., & Chambless, D. L. (1995). Adherence during sessions and homework in cognitive-behavioral group treatment of social phobia. *Behaviour Research and Therapy*, *33*, 573–577.
- Eisler, R. M., Hersen, M., Miller, P. M., & Blanchard, E. B. (1975). Situational determinants of assertive behaviors. *Journal of Consulting and Clinical Psychology*, *43*, 330–340.
- Eisler, R. M., Miller, P., & Hersen, M. (1973). Components of assertive behavior. *Journal of Clinical Psychology*, *29*, 295–299.
- Galassi, J. P., Hollandsworth, J. G., Radecki, C., Gay, M. L., Howe, M. R., & Evans, C. L. (1976). Behavioral performance in the validation of an assertiveness scale. *Behavior Therapy*, *7*, 447–452.
- Glasgow, R. E., & Arkowitz, H. (1975). The behavioral assessment of male and female social competence in dyadic heterosocial interactions. *Behavior Therapy*, *6*, 488–498.
- Glueckauf, R. L., & Quittner, A. L. (1992). Assertiveness training for disabled adults in wheelchairs: Self-report, role-play, and activity pattern outcomes. *Journal of Consulting and Clinical Psychology*, *60*, 419–425.
- Gorecki, P. R., Dickson, A. L., Anderson, H. N., & Jones, G. E. (1981). Relationship between contrived in vivo and role-play assertive behavior. *Journal of Clinical Psychology*, *37*, 104–107.
- Haynes, S. N., & O'Brien (2000). *Principles of behavioral assessment: A functional approach to psychological assessment*. New York: Plenum/Kluwer.

- Hope, D. A., & Heimberg, R. G. (1988). Public and private self-consciousness and social phobia. *Journal of Personality Assessment*, *52*, 626–639.
- Hope, D. A., & Heimberg, R. G. (1993). Social phobia and social anxiety. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders: A step-by-step treatment manual* (2nd ed., pp. 99–136). New York: Guilford Press.
- Hope, D. A., Heimberg, R. G., & Bruch, M. A. (1995). Dismantling cognitive-behavioral group therapy for social phobia. *Behaviour Research and Therapy*, *33*, 637–650.
- Hope, D. A., Heimberg, R. G., Juster, H., & Turk, C. L. (2000). *Managing social anxiety and social phobia*. Albany, NY: Graywind.
- Hope, D. A., & Mindell, J. A. (1994). Global social skill ratings: Measures of social behavior or physical attractiveness? *Behaviour Research and Therapy*, *32*, 463–469.
- Kelly, J. A., St. Lawrence, J. S., & Brasfield, T. L. (1991). Predictors of vulnerability to AIDS risk behavior relapse. *Journal of Consulting and Clinical Psychology*, *59*, 163–166.
- Kern, J. M. (1982). The comparative external and concurrent validity of three role-plays for assessing heterosocial performance. *Behavior Therapy*, *13*, 666–680.
- Kern, J. M. (1991). An evaluation of a novel role-play methodology: The standardized idiographic approach. *Behavior Therapy*, *22*, 13–29.
- Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H. U., Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders among people aged 15–54 in the United States: Results from the National Co-morbidity Survey. *Archives of General Psychiatry*, *51*, 8–19.
- Kirchner, E. P., Kennedy, R. E., & Draguns, J. G. (1979). Assertion and aggression in adult offenders. *Behavior Therapy*, *10*, 452–471.
- Lewinsohn, P. M., Mischel, W., Chaplin, W., & Barton, R. (1980). Social competence and depression: The role of illusory self-perceptions. *Journal of Abnormal Psychology*, *89*, 203–212.
- Liberman, R. P. (1982). Assessment of social skills. *Schizophrenia Bulletin*, *8*, 62–83.
- McFall, R. M. (1982). A review and reformulation of the concept of social skills. *Journal of Behavioral Assessment*, *4*, 1–33.
- Meier, V. J., & Hope, D. A. (1998). Assessment of social skills. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (4th ed., pp. 232–255). Needham Heights, MA: Allyn & Bacon.
- Merluzzi, T. V., & Biever, J. (1987). Role-playing procedures for the behavioral assessment of social skill: A validity study. *Behavioral Assessment*, *9*, 361–377.
- Mueser, K. T., Valenti-Hein, D., & Yarnold, P. R. (1987). Dating-skills groups for the developmentally disabled: Social skills and problem-solving versus relaxation training. *Behavior Modification*, *11*, 200–228.
- Nelson, R. O., Hayes, S. C., Felton, J. L., & Jarrett, R. B. (1985). A comparison of data produced by different behavioral assessment techniques with implications for models of social skills inadequacy. *Behaviour Research and Therapy*, *23*, 1–11.
- Newman, M. G., Hofmann, S. G., Trabert, W., Roth, W. T., & Taylor, C. B. (1994). Does behavioral treatment of social phobia lead to cognitive changes? *Behavior Therapy*, *25*, 503–517.
- Nietzel, M. T., & Bernstein, D. A. (1976). Effects of instructionally mediated demand on the behavioral assessment of assertiveness. *Journal of Consulting and Clinical Psychology*, *44*, 500.
- Romano, J. M., & Bellack, A. S. (1980). Social validation of a component model of assertive behavior. *Journal of Consulting and Clinical Psychology*, *48*, 478–490.

- St. Lawrence, J. S., Kirksey, W. A., & Moore, T. (1983). External validity of role play assessment of assertive behavior. *Journal of Behavioral Assessment*, 5, 25–34.
- Torgrud, L. J., & Holborn, S. W. (1992). Developing externally valid role-play for assessment of social skills: A behavior analytic perspective. *Behavioral Assessment*, 14, 245–277.
- Trower, P. (1980). Situational analysis of the components and processes of behavior of socially skilled and unskilled patients. *Journal of Consulting and Clinical Psychology*, 30, 526–537.
- Trower, P., Bryant, B., Argyle, M., & Marziller, J. (1978). *Social skills and mental health*. Pittsburgh, PA: University of Pittsburgh Press.
- Twentyman, C. T., & McFall, R. M. (1975). Behavioral training of social skills in shy males. *Journal of Consulting and Clinical Psychology*, 43, 384–395.
- Walters, K. S., & Hope, D. A. (1998). Analysis of social behavior in individuals with social phobia and nonanxious participants using a psychobiological model. *Behavior Therapy*, 29, 387–407.
- Wessberg, H. W., Marriotto, M. J., Conger, A.),, Farrell, A. D., & Conger, J. C. (1979). The ecological validity of role plays for assessing heterosocial anxiety and skill of male college students. *Journal of Consulting and Clinical Psychology*, 47, 525–535.
- Woody, S. R., Chambless, D. L., & Glass, C. R. (1997). Self-focused attention in the treatment of social phobia. *Behaviour Research and Therapy*, 35, 117–129.