CHAPTER 3

Methods of Interpersonal Skill Assessment

Brian H. Spitzberg
San Diego State University

THE KEN OF INTERPERSONAL SKILLS ASSESSMENT

Few characteristics are more important to the everyday quality of life as the skill with which interpersonal communication is negotiated. Yet few concepts are as difficult to define and assess as interpersonal skill. This chapter examines issues associated with the assessment of social interaction and communication skills. It proceeds by first considering the importance of such skills. It continues by making several key distinctions in terminology and concepts relevant to a review of assessments. Next, a synoptic overview of historical eras is provided, which is intended to situate current debates about assessment concerns. These current debates are presaged in a discussion of several ideological tensions often overlooked in the examination of skills assessment. This discussion gives way to a characterization of several trends in assessment that have emerged since the last major reviews were written before the 1990s (e.g., Bellack, 1979, 1983; Curran, 1979a, 1979b; Curran, Farrell, & Grunberger, 1984; Curran & Mariotto, 1980; Kolko & Milan, 1985b; McFall, 1982; Spitzberg, 1988, 1989; Spitzberg & Cupach, 1989) in light of those in the interim (e.g., Hargie, 1997; Inderbitzen, 1994; Matson, Sevin, & Box, 1995; Rubin, 1994; Spitzberg, 1994b; Trower, 1995). Then the chapter explicates a conceptual heuristic, the adapted Behavioral Assessment Grid (BAG, Cone, 1978; Spitzberg, 1988; Spitzberg & Cupach, 1989) for analyzing assessment methods. With this conceptual vocabulary established, a number of omnibus measures and methods are briefly reviewed (leaving the more specific contextual and skill-specific assessments to the appropriate later chapters of this text). The chapter concludes with a consideration of key decision points any scholar or practitioner should consider in developing an assessment and some of the implications of these decisions.
THE IMPORTANCE OF INTERPERSONAL SKILLS

There is abundant evidence that competence in interpersonal skills is vital to psychological, emotional and physical health (Spitzberg & Cupach, in press). Previous review (Spitzberg & Cupach, 1989) has concluded that interpersonal competence is empirically associated with the etiology and prognosis of mental disorders (e.g., Monti et al., 1984; Monti & Fingeret, 1987), anxiety (e.g., Conger, Wallander, Mariotto, & Ward, 1980; Fydrich, Chambless, Perry, Buergener, & Beazley, 1998), stress (Herzberg et al., 1998), cardiovascular disease (Ewart, Taylor, Kraemer, & Agras, 1991), loneliness (Spitzberg & Hurt, 1987b), academic success (Rubin, Rubin, & Jordan, 1997), juvenile delinquency (Renwick & Emler, 1991), drug abuse (Twentyman et al., 1982), dysphoria and depression (Segrin, 1998). Negative, compared with positive, communication has been associated with much stronger influence on marriage (Gottman, 1994), relational satisfaction (Spitzberg, Canary, & Cupach, 1994) and psychological well-being (Spitzberg & Cupach, in press). Interpersonal and communication skills, broadly defined, are consistently ranked as among the top two or three competencies that organizations require of their employees (e.g., O’Neil, Allred, & Baker, 1997).

Several other findings suggest, at minimum, an indirect or mediational role of interpersonal skills (Spitzberg & Cupach, in press). For example, House, Landis, and Umberson (1988) summarized epidemiological studies to find a consistent effect of social integration on mortality rates. Many of these studies found these effects even after controlling for drug use, smoking, obesity, and health care practices. Another indication of the indirect impact of interpersonal skills is the “marriage benefit.” Data show that married people, compared with single or divorced people, have lower suicide rates, imprisonment rates, psychiatric diagnoses, and mortality rates (Argyle, 1991).

Interpersonal interaction is the sine qua non of marriage, family, and social integration. It can be accepted as axiomatic that the more interpersonally skilled a person is, the more likely it is that person will successfully negotiate satisfying marriage, family, and extended networks of social relationships. It is further accepted as axiomatic that higher levels of interpersonal skills either directly or indirectly facilitate significantly higher levels of psychological, emotional, and physical well-being (Segrin & Flora, 2000).

Were everyone interpersonally competent, the alarming findings regarding interpersonal skills and well-being would hardly be a major concern. Important social competencies evade large proportions of the population, however (see reviews by Spitzberg, 1987; Spitzberg & Cupach, 1989). Estimates indicate that at least 7% to 10% of the population is socially inadequate (Curran, Miller, Zwick, Monti, & Stout, 1980; Hecht & Wittchen, 1988), although some would estimate the rate at closer to 25% (Bryant, Trower, Yardley, Urbiet, & Letemendia, 1976; Vangelisti & Daly, 1989). Such inadequacies may explain why as much as one fifth to one quarter of the population suffer from loneliness (Perlman & Peplau, 1982), anxiety, or shyness (Richmond & McCroskey, 1985; Zimbardo, 1977). In summary, substantial proportions of the population experience difficulties with their social interaction and interpersonal relationships, and these difficulties are associated with myriad psychological, emotional, and physical problems (Segrin, 1998; Segrin & Flora, 2000). It is no small thing, therefore, to inquire as to the state of social interaction and communication skills assessment.
3. METHODS OF INTERPERSONAL SKILL ASSESSMENT

TERMS AND DISTINCTIONS

Skills

Despite extensive efforts “there is little consensus in the field regarding the definition of social skills” (Bedell & Lennox, 1997). Skills are defined here as intentionally repeatable, goal-directed behaviors and behavior sequences (Spitzberg & Cupach, 1984). Skills are the actual behaviors manifested in the attempt to accomplish some goal. These behaviors are repeatable, more or less, on demand. “Communication” and “social interaction” will be taken here as interchangeable, even though there are communication situations that can be arguably considered nonsocial or noninteractive. In reference to social interaction, therefore, such skills presuppose interdependent goals, goals that can only be accomplished through symbolic interaction with others. Furthermore, such skills must be intentionally repeatable. Almost anyone may be able to introduce himself or herself to an attractive stranger at some point. But to be able to do so at will implies that the person has a skill of initial interaction.

The definition of interpersonal skills as behavioral captures a particular view of skills. Many authorities concur that skills should be conceptualized and assessed at the behavioral level (e.g., Bellack & Hersen, 1978; Curran, 1979a, 1979b; Hargie, 1997). Others, in contrast, have made articulate cases for including social cognitive and perceptual abilities (e.g., Burleson, 1987) or intrapersonal production abilities (e.g., Greene, 1984, 1994, 1997). The approach taken here is that it is conceptually helpful to separate the motivational and knowledge factors from skill factors, and to further separate those factors that account for the production of behavior from competence, which is the evaluation of the quality of performed behavior (McFall, 1982; Spitzberg & Cupach, 1984). To a large extent, this is the distinction Trower (1984) recommended between social skill (i.e., motivation and knowledge) and social skills (i.e., the manifest behaviors that are a product of motivation and knowledge).

Clearly, skills cannot be produced without the skill, but the two terms imply distinct content domains of assessment. This chapter primarily concerns assessments that focus on the objective or subjective representation of behavioral referents, although passing mention is made of the motivation and knowledge domains, given their close relevance to interpersonal skills as they are defined here.

Skills, therefore, are generally thought to be manifestations of some underlying ability, which is a capacity for action. This capacity is typically conceptualized as a function of numerous motivation (e.g., confidence, goals, reinforcement potential, etc.) and knowledge (e.g., content and procedural knowledge, familiarity, etc.) components.

This discussion of skills and ability has foregone any mention of success in goal accomplishment. This is an extraordinarily complex issue that belies what otherwise seems like a straightforward concept. Consider a relatively standard objective assessment, the eye test. In a typical eye test, subjects are asked to read a series of sequentially smaller rows of alphabetic letters to determine visual skill. Underlying this skill is not only visual acuity (i.e., the eye’s ability to receive visual input at various ranges, in various colors, with varying degrees of discrimination of line and form) and cognitive interpretation (i.e., the ability to distinguish concepts such as two dimensions vs. three, solid vs. nonsolid, etc.), but also symbol recognition (i.e., the ability to know and recognize alphabetic and numeric symbols). Indeed, technically speaking, the vision test also involves the skill of communicating the end
product of these underlying abilities (i.e., actually saying the “words” “E, W, O, 2, F” and so on, in correspondence with one’s interpretations). At first blush, this seems an objective assessment of skill. However, consider some of the presumptions built into the assessment. Were it a preliterate culture, nonverbal iconic symbols would take on much greater social relevance than linguistic symbols. For people with certain mental disabilities, such a linguistic basis for assessing reading may be relatively meaningless. Furthermore, why is the ability to read symbols important? Perhaps because “literate” societies have deeply embedded values favoring the ability to read, which in turn depends on alphabetic recognition (e.g., traffic signs, advertising, instructions, bureaucratic forms, etc.). So recognizing linguistic and numeric symbols is taken as an assessment stimulus of choice because society values the importance of those particular symbols as indicators of social competence. The eye exam ultimately assesses several underlying abilities, and collectively the behavioral product of these abilities is taken as a proxy of a person’s visual skill. The point is that even an assessment as objective as the eye exam is subtly imbued with a host of subjective decisions. The eye exam seems uncontroversial because it is employed in a societal context that reveals its relevance in the normative fabric of everyday interaction. Remove it from that context, and suddenly its relevance to a concept of competence is problematic.

The eye test is relevant only in a societal context, and the context itself is a highly multifaceted concept. The term context represents at least five clusters of meaning, each of which is important to the assessment of skills. Context can be understood in terms of culture, time, relationship, situation, and function. Culture entails the sets of behaviors, beliefs, values, and linguistic patterns that are relatively enduring over time and generation within a group (Gudykunst & Ting-Toomey, 1988). Skills valued in one culture are clearly not necessarily the skills valued in another (e.g., Hecht, Collier, & Ribeau, 1993; Nicotera, 1997). Time refers to the sequential nature of skills (e.g., questions tend to precede answers), the use of time in context (e.g., amount of time spent speaking), and the frame of time across which skills are assessed (e.g., state vs. trait). Interpersonal skills tend to be sequentially organized (Psathas, 1990). Use of time within an episode of interaction is predictive of perceived competence (Dillard & Spitzberg, 1984). In addition, skills manifest in a given episode of interaction often bear only minimal relationship to skills assessed over time (Cupach & Spitzberg, 1983; Spitzberg, 1987, 1990; Spitzberg & Hurt, 1987b). Context can also be viewed as the type of relationship between interacants. Typical relationships include kinship (e.g., parent–child, sibling, etc.), intimate (e.g., dating, married, etc.), social (e.g., acquaintance, friend, etc.), instrumental (e.g., superior–subordinate, colleague, etc.), or more generic (e.g., stranger) connections. Skills valued in one type of relationship are not necessarily valued in another (Hecht, Serena, & Spitzberg, 1984; Knapp, Ellis, & Williams, 1980). Context also takes the form of the physical or social situation in which interaction occurs. Situations vary in a variety of characteristics, including the more sensorial ways (e.g., temperature, spatial arrangements, etc.) and the more social ways (e.g., formal–informal, active–inactive, etc.). Skills valued in some situations, such as a formal interview, may not be valued in other situations, such as an informal chat over coffee (Argyle, Furnham, & Graham, 1981; Pavitt, 1989). Finally, contexts vary according to the function being served by the interaction (Burleson, 1987). Communication does rather than just is. As such, skills valued for one function (e.g., assertion) may not be valued in the pursuit of another function (e.g., affection).

Observing that skills are, or are not, valued in given contexts suggests that skills alone are rarely the sole issue when assessing communication. Instead, skills are
typically only important in society to the extent they are considered competent or incompetent. The competence of skills, it turns out, is often a much more complicated concept than the skill itself.

**Competence**

Competence can be viewed as an evaluative judgment of the quality of a skill (McFall, 1982; Spitzberg & Cupach, 1984). The majority of contemporary competence literature has focused on appropriateness and effectiveness as core criteria. The importance of these criteria is clarified by a brief consideration of the alternative criteria that are sometimes forwarded (Spitzberg, 1993, 1994a, 2000; Spitzberg & Cupach, 1984, 1989, in press).

**Criteria of Competence**

**Dialogical Criteria.** Dialogue is defined by such characteristics as “coordination (or cooperation), coherence, reciprocity and mutuality (e.g., with regard to moral commitments)” (Linell, 1998, p. 14). Dialogue emphasizes skills such as empathy, confirmation, relaxed readiness, perspective reflection, meta-communication, congruence, humor, present orientation, genuineness, and egalitarianism (Kristiansen & Bloch-Poulson, 2000; Pearce & Pearce, 2000). Such approaches are related to critical perspectives attempting to construct an ethical system of social discourse (e.g., Habermas, 1970; see Burleson & Kline, 1979; Penman, 1992).

**Clarity.** Clarity is one of the most intuitive or lay notions of competence (McCroskey, 1984; Powers & Spitzberg, 1986) and reflects a natural language perspective in which language, used properly, is thought to have the capacity for reflecting an observational world (Clark & Paivio, 1989). It is typified by such statements as “Why can’t you just say what you mean” or “Just be clear.” Clarity can be viewed in relatively objective terms (e.g., readability indexes) or in somewhat more subjective senses (e.g., code elaborateness).

**Understanding.** An implicit conjunction with clarity, understanding is also a common criterion of competence. Typified by statements of “You just don’t understand what I’m saying,” and “We need to understand each other better,” this criterion is often confounded with clarity. Clarity is a characteristic of expressive behavior. Words can be more or less clear based on such things as definitional complexity, rarity, or contextual specificity. Understanding, in contrast, is a mentalistic notion. Independent of mode of expression, to what extent do interlocutors comprehend each others’ intended meanings? Specifying the nature of understanding is itself a controversial subject (e.g., Ickes, 1997; Kenny, 1994).

**Efficiency.** Efficiency refers to the notion that accomplishing a goal can involve less or more effortful and resource-intensive activities. A person is more competent to the extent that less resource-intensive, complex, or effortful means are employed to achieve a given goal (Kellermann & Park, 2001). Efficiency is typified by statements such as “Well why didn’t you just say so in the first place?” and “Why are you beating around the bush?”
**Satisfaction.** A person who accomplished preferred outcomes through interaction is likely to experience a sense of positive affect, or satisfaction (Spitzberg & Hecht, 1984). Satisfaction is typified by statements such as, “I really enjoyed talking with you” and “How do you feel about your interview?” Even when interaction is normatively unpleasant, such as conflicts, a person can be more satisfied with some responses relative to other responses.

**Effectiveness.** Effectiveness is the extent to which preferred outcomes are achieved. Effectiveness is obviously related to efficiency and satisfaction, both of which employ effectiveness as one of their definitional components. Effectiveness is perhaps the most elemental representation of the functional aspect of communication. Communication is enacted to accomplish something, and the extent to which this something is achieved provides a measure of the competence of that communication (Parks, 1985).

**Appropriateness.** A further criterion often attributed to communication is appropriateness, which is the extent to which behavior meets the standards of legitimacy or acceptability in a context (Larson, Backlund, Redmond, & Barbour, 1978). Appropriateness has occasionally been defined as the extent to which behavior conforms to existing contextual rules, but this is an unnecessarily delimiting construction. Sometimes the most competent behavior is to alter existing rules or establish new rules. Thus, appropriateness is better viewed in terms of behavior that accords with the extant (rather than existing) rules of an interpersonal context (Spitzberg, 2000).

Appropriateness and effectiveness represent the most general, encompassing, and conceptually useful criteria for competence (Spitzberg, 1983, 2000; Spitzberg & Cupach, 1984). Clarity and understanding are only important to the extent that interactional goals are achieved. Efficiency adds a value judgment that the quickest or least “expensive” path is always preferable to one that may take more effort but end up being more rewarding. This value judgment seems unnecessary to the primary concern of most interactants to be appropriate and effective. Satisfaction runs afoul of such distortions as those who feel good about their coercive behavior, or a performance that was viewed as inept by everyone else in the encounter except the performer. Relying on either appropriateness or effectiveness alone leads to similar objections. However, combining appropriateness and effectiveness provides a framework that most competence theorists accept as generally viable. Competence, according to the dual criteria of appropriateness and effectiveness, is the extent to which an interactant achieves preferred outcomes in a manner that upholds the emergent standards of legitimacy of those judging the interaction.

**SUMMARY**

Communication and social interaction skills are viewed here as a set of behaviors and behavior sequences. “Asking a question” or “making eye contact” are examples of interpersonal skills. Whether these skills were enacted in a way that was successful, satisfying, appropriate, clear, and so forth is a matter of quality or competence. The ultimate purpose of assessing communication and social interaction skills rarely turns on the mere ability to enact skills. Instead, the purpose of most assessment efforts is to obtain a normative reference point on an implicit or explicit continuum of social competence.
This chapter is about the assessment of both skills and competence, because the two are so often inextricably intertwined. The normative nature of the connection between skills and competence is in evidence throughout history. Therefore, a brief historical sketch of conceptions of communication and social interaction skills assists in situating current assessment practices.

A HISTORICAL OVERVIEW

Different cultures and eras produce relatively distinct epistemes of competence, across and within which both the content and criteria of competent interaction vary. For example, the “upper crust” of society may have one set of standards and the “lower crust” may have another. Recognizing the limitations and Western biases implied by written histories, several chronologically ordered epistemes can be identified (see Phillips, 1985; Spitzberg & Duran, 1993; Wine, 1981). The ancient rhetorical tradition, given classical Greece and Rome’s reliance on public oratory in matters of state and judiciary, focused on the mastery of oral argument in the pursuit of successful persuasion. The sophists presupposed that communication skills were teachable and that those with better skills would benefit proportionately from communication skill. Soon, Aristotle’s rhetoric, defined as the study of the available means of persuasion, would dominate discussion until the 18th century.

Before and during the Renaissance, as the Black Death receded and the remaining resources were available to smaller numbers of people, divisions of wealth permitted a broader and more variegated status hierarchy. But as wealth tends to motivate actions to preserve that wealth, rhetorics evolved that erected social barriers to parallel the resource barriers. The politics of holding court involved elaborate social codes of conduct (Jeanneret, 1991; Menache, 1990). By the end of the 17th century, influential books by Castiglione, Guazzo, and Della Casa articulated these social codes as maps to social mobility, even if such mobility were practically reserved for those already in the higher reaches of the status hierarchy (Goldsmith, 1988).

Partly concomitant with the politics of courtship was the episteme of manners and etiquette in which books were written to serve as arbiter elegantiarum for an age. Such concerns were presaged by the elocutionist movement of 19th and early 20th centuries rhetorics in which very specific stylized skills and movements were taught as the proper expressions of certain emotions and intentions (Austin, 1966; Sheridan, 1762). Ewbank’s (1987) examination of American etiquette books of the early 20th century concludes that behavior books for children, written largely by clergy, focused on moral principles of action, whereas adult books reflected imported notions of European aristocracy. It is not entirely coincidental that in the first half of this century conceptualizations of social intelligence, relying on racially and gender-biased notions of cognitive intelligence (e.g., Doll, 1935, 1953), were at their zenith.

The importance of this selective and abstracted historical review of the rhetoric(s) of competence is to illustrate that competence is culturally, chronologically, and contextually evolutionary. The concept of “skill” is inherently relative in importance and reflective of prevailing ideologies.

Some Ideological Dialectics and Tensions in Assessment

If competence and interpersonal skills are intrinsically intertwined with ideology, it is important to identify the tensions created when one value is given preference and another marginalized. There are many such contradictions and dialectics
Elaboration Versus Conciseness. A vision of expressiveness, of self-disclosure, of style and flourish would emphasize competence as elaboration. Furthermore, societies that value clear status differences in their hierarchies might value elaborate codes of interaction, in which, as Bernstein (1986) noted, the upper-crust groups master a more complex and differentiated set of symbolic resources and rules of use than more pedestrian native interactants. An efficiency ethic prefers a more mechanical or industrial criterion, in which unnecessary words and actions are removed like chaff from the wheat.

Understanding Versus Feeling Understood. The ethic of understanding claims that communication functions best when mutual understanding is achieved. The understanding ethic is reflective of a conduit metaphor of communication, within which an entity or state in one interlocutor's mind is transferred into the other interlocutor's mind. The precision with which this transfer occurs, and the extent to which the entity is reproduced in the other's mind without distortion, are measures of the communication skill of the interactants. In contrast to being understood, much of the rhetoric of the 1960s and 1970s celebrated the importance of feeling good. Indeed, there is a fair amount of research to suggest that feeling understood is strongly associated with relational satisfaction (Cahn, 1990). There is relatively little research suggesting that actual understanding and accuracy is related to relational satisfaction (Spitzberg, 1993).

Honesty−Assertiveness Versus Equivocation−Politeness. An ethic began to emerge in the 1960s, particularly in American therapeutic literatures, that emphasized assertiveness, honest expression, and disclosure of feelings. Genuineness of self-expression, openness, and directness were highly valued as pathways to self-discovery and authentic intimacy (see Bochner, 1982; Parks, 1982; Spitzberg, 1993). In contrast, others point out the universal pragmatic of politeness (Brown & Levinson, 1987) and the essential nature of equivocal communication in preserving the delicate machinations of everyday face support (e.g., Chovil, 1994). Deception can be competent and can be based on altruistic motives, but it is normatively viewed as incompetent (O'Hair & Cody, 1994).

Effectiveness Versus Appropriateness. Despite the virtually univocal acceptance of these dual criteria among contemporary competence theorists, there are frequent tensions involved in achieving both criteria simultaneously (Spitzberg, 1993, 1994a, 1994c). Conflict encounters, for example, represent contexts in which the effectiveness of parties is generally viewed as mutually exclusive, and therefore, lines of appropriate action are highly restrictive (Spitzberg et al., 1994). To be appropriate is to occasionally forego one's optimum efficacy, and to be entirely effective at times is to jeopardize one's appropriateness in the eyes of others.

Summary

The history of competence reveals that there are ebbs and flows in the evolution of societal conceptions of "skilled" communication. For example, competence defined in terms of empathy, cooperation, dialogue, understanding, and appropriateness is
more in line with traditional feminine stereotypes (Spitzberg, 1994b). In contrast, competence defined in terms of assertiveness, efficiency, and effectiveness is more in line with traditional masculine stereotypes. Such definitional criteria are therefore not without their direct implications for the type of society they envision and the type of society in which such criteria currently reside. Any discussion of competence assessment that is unaware of such implications is shortsighted and ultimately at risk of being misdirected in politically pernicious or arbitrary ways.

METHODS

Current Trends

The 1970s and 1980s represented a very active time in the scholarship on social skills. The concept was relatively recent in its articulation (Argyle, 1969). When it began to be articulated by respected scientists as an alternative model for mental health, and thus intervention (Phillips, 1978; Trower, Bryant, & Argyle, 1978), it quickly captivated the scholarly imagination. By the mid-1980s, there were manifold reviews and discussions of challenging or intractable issues involved in the assessment of social skills (see, for example, Bellack, 1979, 1983; Curran, 1979a, 1979b; Curran & Mariotto, 1980; Curran et al., 1984; McFall, 1982; Spitzberg, 1987, 1994a, 1994c; Spitzberg & Cupach, 1989). It seems surprising, however, that more than a decade later there are only a handful of new measures to add to the list. There is a sense in which social skills and communication competence have receded from the forefront of scholarly interests. To the extent to which this is more than mere appearance, at least three explanations are evident, all having to do with social and communication skills research moving to particular pastures rather than disappearing altogether from the scholarly landscape.

First, there seems to be a trend toward specialized uses of the social skills and competence concepts. These concepts once were envisioned as integrative metaphors for the entire field of interpersonal communication. More recently, they seem more relegated to three fields of endeavor: therapeutic (including marital), intercultural, and instructional contexts.

Second, interest in social and communication skills assessment may have simply moved into more specialized arenas. There are still vibrant veins of research in clinical and counseling literatures on social skills both as causes of social and psychological problems and as a potential source for effective intervention. These literatures appear to have become very specific applied tributaries of what was once a more general mainstream of social skills research. Social and interpersonal skills are being examined in relation to such specific arenas as loneliness, depression, dysphoria (Segrin, 1998), gender (Bruch, Berko, & Haase, 1998), delinquency (Gaffney, 1984), health care delivery (Ravert, Williams, & Fosbinder, 1997), heterosocial interaction (Kolko & Milan, 1985a; Perri, Richards, & Goodrich, 1978; Wallander, Conger, & Conger, 1985) and marital interaction (e.g., Gottman, 1994). The trend is toward specialized contextual applications of the assessment of social interaction skills, rather than toward general theories or assessment approaches.

Third, social skills assessment has moved into more interdisciplinary tributaries. For example, competence is a dominant theme throughout the educational context. It is difficult to pick up an undergraduate textbook in communication that does not espouse a competence framework. At a more scholarly level, however, there is
considerable research being conducted on various competencies, including commu-
nication competencies, in terms of educational achievement and curricula. It is in the
educational literature that general measurement and comparative validation studies
are still being conducted (e.g., Demaray et al., 1995; Flanagan, Alfonso, Primavera,
Povall, & Higgins, 1996; Smit & van der Molen, 1996) and general conceptual frame-
works for assessment (e.g., Sheridan, Hungelmann, & Maughan, 1999) still seem to
be active areas of scholarly endeavor. There also seems to be considerable inter-
communication skills in corporations and their assessment centers (e.g., O’Neil,
Allred, & Dennis, 1997).

Despite the apparent waning of excitement over social and communication skills
at the scholarly level, researchers are still conducting significant and substantive
work. But as indicated above, the literature and relevant assessments are scattered
far and wide across the scholarly terrain. To facilitate a review of assessments and
the key issues involved, an organizational heuristic, the Behavioral Assessment Grid,
is explicated next.

Overview of the Adapted Behavioral Assessment Grid

Traditional reviews of social skills assessment tend to classify methods and measures
into types such as direct behavioral observation, behavior-rating scales, interviewing
techniques, sociometric methods, objective self-reports, role-play methods, and so
forth (e.g., Merrell, 1994). Such approaches gloss over the important conceptual di-
mensions underlying and differentiating these diverse methodologies. It is important,
therefore, to attend to those underlying dimensions that organize and differentiate
assessments. Issues such as what is being assessed, who is doing the assessing, the
level of abstraction of the referents being assessed, and the level of generalization im-
licit in the assessment represent fundamental distinctions among the various types
of assessment. The Behavioral Assessment Grid (BAG) is a heuristic model that assists in
identifying these dimensions.

The Behavioral Assessment Grid (BAG) was originally conceived by Cone (1978)
to organize discussion about behavioral assessment issues. It is adapted here and
elsewhere (Spitzberg & Cupach, 1989) to provide a useful vocabulary for analyzing
various options and issues related to skills assessment (see Fig. 3.1). Because many
assessments ultimately represent hybrids of various forms, the categories of the BAG
are not intended to be mutually exclusive.

The BAG is a matrix defined by three axes: assessment referent, method, and uni-
verses of generalization. Assessment referent is the behavioral content or domain of
reference (e.g., cognition vs. behavior). Method refers to the techniques of generating
assessment referents (e.g., role-play vs. self-report). Universes of generalization refer
to the domains across which assessments are intended to apply (e.g., time vs. context).

Assessment Referent. Assessment of social interaction skills must refer to some
aspect of behavior. But behavior can be referred to in at least three distinguish-
able domains: motivation, knowledge, and skill. Any assessment of social or inter-
action skills could refer to motivational characteristics, including issues of arousal,
anxiety, nervousness, apprehension, interest, goals, desires, needs, effort, or val-
ues. Such references can be made to internal states (e.g., “I really want to make a
good impression when I meet strangers”) or to behavioral activities presumed to
result from such internal states (e.g., “I really try to make a good impression when
I meet strangers”). Assessment of skills can also examine knowledge, or cognitive,
3. METHODS OF INTERPERSONAL SKILL ASSESSMENT

Methods. Methods refer to the ways in which referents are either stimulated or presented for assessment. A role-play method presents a description of a situation to elicit overt response behavior from the subject. These overt responses are subsequently rated, coded, or otherwise evaluated in regard to their competence. In contrast, questionnaires present subjects with items to rate in reference to their own, or someone else's (e.g., their marital partner's) skills, competence, or both. These methods can be viewed in terms of their degree of removal from the referent being assessed. Indirect methods tend to involve no actual referent skill being manifested in the assessment context. Direct methods, in contrast, elicit performance of overt skills, which are then further assessed in the process.

Indirect methods include projective, interview, self-reference questionnaires, and other-reference questionnaires. Projective methods, such as inkblot techniques or sentence-completion tasks, present ambiguous stimuli to a subject, to which the subject reacts or provides an interpretation, and these reactions are then interpreted according to some scheme designed to assess the subject's social competence (e.g., Helper, 1970).

Interview methods refer to verbally presented questions about the subject's social interaction (e.g., Brugha et al., 1987; Gurland et al., 1972; Hecht & Wittchen, 1988; Monti, 1983). Interview methods are typically employed when there is background or personal information that would be difficult to obtain through behavioral observation.
or without the subject’s individual familiarity with the referents. Self-reference questionnaires refer to the presentation of items (e.g., “I am a good communicator”) the subject applies to himself or herself. Items can vary across any of the universes of generalization to be discussed below. Other-reference questionnaires refer to items (e.g., “My partner is a good communicator”) the subject applies to another person. It is presumed that the respondent has sufficient knowledge of the other person to make the judgments called for by the assessment. Other-reference questionnaires can be scaled in all the same ways as self-reference questionnaires.

Direct methods elicit some sample of behavior that serves as the referent for further assessment and evaluation. Self-monitoring is a technique employed in many therapies in which a person is instructed to code each time she or he engages in some target behavior(s).

In role-play methods, a subject is presented with one or more simulated versions of an interaction situation. The situations are generally developed and selected on the basis of their relevance to the particular areas of skill of concern to the assessors. The subject’s performance is subsequently rated by self, by the confederate, or observers, and the recorded behavior from the scenarios can subsequently be coded and rated (e.g., Bellack, Hersen, & Lamparski, 1979; Bellack, Hersen, & Turner, 1978; Kern, 1982; Kern, Miller, & Eggers, 1983; McNamara & Blumer, 1982; St. Lawrence, Hughes, Goff, & Palmer, 1983).

Naturalistic assessments are situations introduced to subjects as if they were real but involve some degree of manipulation on the part of the assessors (e.g., Bellack, Hersen, & Turner, 1979). These methods tend to be employed when the subjects are in a relatively contained environment, such as patients in psychiatric hospitals or subjects in a waiting room. In vivo assessments involve assessments of people’s behavior in truly natural, unmanipulated contexts (e.g., Snyder & Shanks, 1982). For example, recordings of telephone conversations, courtroom interactions, doctor–patient interviews, classrooms, and even unobtrusively observed interactions in parks or restaurants, all represent examples of in vivo contexts (Psathas, 1990). Finally, task-based or objective criterion assessments represent situations in which a subject’s achievement of some particular outcome provides a measure of that person’s skill, such as compliance-gaining success, accuracy of giving directions or intended meanings (Burleson & Denton, 1992, 1997; Powers & Spitzberg, 1986; Rubin, 1982).

Universes of Generalization. Symbols are not the things to which they refer; that is, the map is not the territory. Every assessment is an abstraction of its referent. As an abstraction, every assessment represents a degree of generalization from that which it is derived. A subject’s glance is generalized to a measure of the subject’s eye contact, the eye contact is generalized to the subject’s confidence or attentiveness, which in turn are generalized to the subject’s autonomy or empathy, and ultimately, the person’s interpersonal competence (Spitzberg & Cupach, in press).

There are at least three domains of generalization: external, internal, and observer. External domains are concerned with the extent to which generalizations are sought across distinct methods, contexts, or times. These are external to the assessment itself in the sense that entirely distinct assessments or referents are being compared. Internal generalizations involve comparisons within the content of a given assessment method. For example, Riggio’s (1986) measure originally conceptualized seven factors (e.g., emotional expressivity, social sensitivity, etc.). Internal generalization asks the extent to which these are intercorrelated dimensions, and equally requisite to the entire conceptualization of interpersonal competence.
Finally, observer generalization concerns issues of the comparability of raters, coders, observers, and subjects.

External universes of generalization are important in establishing various types of validity of an assessment. For example, studies have examined the validity of role-play methods by comparing the social competence of known groups across role-play, self-reference questionnaires, and interview or in vivo methods (e.g., Kern, 1982; Kern et al., 1983; McNamara & Blumer, 1982; Monti et al., 1983; Smit & van der Molen, 1996; St. Lawrence et al., 1983). Context or setting generalization involves the extent to which an assessment applies across similar or distinct types of situation. Does a measure of negotiation competence generalize to other conflict situations, much less to heterosocial situations? Generalization across contexts necessarily assumes generalization across time. Nevertheless, these are conceptually distinguishable universes. An assessment concerned with nursing competence can refer to nursing contexts of the last 2 weeks or the last 2 years.

Internal universes of generalization concern many of the standard psychometric concerns of reliability and content validity. Dimension–factor generalization refers to whether the content facets of a given assessment generalize to each other. Generalization across items can take two forms: internal reliability and item abstraction. Item abstraction has been the subject of considerable debate and is often referred to as the molar–molecular issue (Caballo & Buela, 1988; Dillard & Spitzberg, 1984; Royce, 1982; Spitzberg & Cupach, 1989). Molecular items are low-abstraction references to behavior. Molecular items tend to refer to relatively discrete, observable, and objectively definable behaviors. Behaviors such as pauses, filled pauses, gestures, eye gaze, body lean, asking of questions, smiles, talk time, fidgets, interruptions, and so forth are considered relatively molecular. Some scholars have conceptualized “midi,” “mezzo,” or “intermediate” level constructs, such as “assertion,” or “humor” (e.g., Farrell, Rabinowitz, Wallander, & Curran, 1985; Monti et al., 1984; Spitzberg & Cupach, in press). Molar items, in contrast, are relatively subjective and evaluative and tend to involve high-level inferences. Items such as “S was a good communicator” reflect a more molar type. The issue of generalization across levels of abstraction concerns the extent to which it is legitimate to sum across such items.

The domain of observer generalization is the extent to which assessments are comparable across raters, coders, observers, and subjects. Any time a coding or rating system is applied by multiple assessors, it is common to ascertain the reliability or correspondence of these assessors. According to traditional psychometrics, a measure cannot be valid if it is unreliable. A coding system that codes for assertive statements cannot be considered a valid assessment of a person’s assertiveness if no two coders see the same behaviors as fitting into the same coding categories. In contrast, it is open to debate as to whether different observers should evaluate competence and skills similarly. Finally, generalization across subject–actor concerns whether a measure applies to many types of subject (e.g., children, adolescents, adults).

Summary. The BAG is not entirely comprehensive nor are its dimensions entirely mutually exclusive, but it does provide a useful working vocabulary for the analysis of most assessments of interaction skill. The matrix formed by the intersection of these axes suggests ways of categorizing assessments, as well as revealing relatively empty cells in which assessment efforts have been slighted. The axes of the BAG lay the conceptual groundwork for review of some of the more prominent assessments of social interaction skills, as well as the subsequent discussion of alternative types of assessment and assessment problems remaining to be resolved.
There are hundreds of communication and social interaction assessments. A comprehensive review is impractical, and a selective review is offered instead. Assessments were excluded if they were (a) developed specifically for a particular research project (e.g., Smit & van der Molen, 1996; Sanford, 1998), (b) unpublished (e.g., Kelly & Chase, 1978), (c) designed for very narrow conceptions of skill (e.g., Ravert et al., 1997; Sharpe, Connell, & Gallant, 1995), context (e.g., Gruppen et al., 1997; Ralph & Thorne, 1993), or population (e.g., see Table 4.5 of Spitzberg & Cupach, 1989), (d) largely ignored in application over the last decade, (e) ambiguous in extent of overlap with interaction skills (e.g., Cegala, Savage, Brunner, & Conrad, 1982), or (f) developed with dimensions extraneous to interaction (e.g., McCroskey & McCroskey, 1988; Rubin, 1982, 1985). What follows is a highly selective tour highlighting some of the trees of the assessment forest. Out of the assessments that were not excluded on the basis of the criteria above, assessments were included if they (a) represented a relatively omnibus measure of communication skill or competence and (b) were published in more than one study.

Furthermore, even though the emphasis of this review is on interpersonal skills, it is important to point out the availability of assessments in the motivation and knowledge domains. As indicated, motivation and knowledge are the underlying abilities or capacities that give rise to interpersonal behavior, and therefore, their relevance to interpersonal skills is clear. There are already excellent reviews available for the motivation domain. Whereas there are few reviews of the knowledge domain of assessment relevant to interpersonal skills, there are numerous assessments that are arguably germane. Therefore, a brief note regarding available measures of motivation and knowledge is provided next.

**Motivation Assessments**

The anxiety component of communication motivation is one of the most established in terms of assessment (Spitzberg & Cupach, 1989; Patterson & Ritts, 1997). There are hundreds of measures of anxiety, nervousness, apprehension, unwillingness to communicate, and shyness. Some of these measures are largely behavioral measures of manifest behavioral anxiety such as fidgeting, filled pauses, and avoidance of eye contact (Mulac & Wiemann, 1997). Others are based on the assumption that anxiety is reliably indicated by physiological arousal (Beatty & Dobos, 1997). Most, however, assume that people are fully cognizant of their nervousness and are able to reliably and validly report their own level of communication apprehension (McCroskey, 1997).

There are far fewer assessments of more positive communicator motivation. Positive communication motivation can be conceptualized in a variety of ways (Zorn, 1993), including self-efficacy (Moe & Zeiss, 1982), sensation-seeking (e.g., Zuckerman, 1994), pursuing and distancing (Bernstein, Santelli, Alter-Reid, & Androsiglio, 1985), extroversion and talkativeness (e.g., McCroskey & Richmond, 1995; Wheeless, Frymier, & Thompson, 1992), motives (Rubin, Perse, & Barbato, 1988), attentiveness (Norton & Pettigrew, 1979), communication involvement (Cegala, 1981), and simply as motivation to communicate competently (Spitzberg & Hecht, 1984).

**Knowledge Assessments**

There have been several measures developed to assess the cognitive or knowledge-based component of communication and social interaction skills. Martin and
Rubin (1995) developed a 12-item measure to tap three components of cognitive flexibility—awareness of options, willingness to adapt, and self-efficacy in adapting—using a Likert-type response scale. It has been studied in relation to interaction involvement, self-monitoring, interpersonal communication competence, rigidity, unwillingness to communicate (Martin & Rubin, 1995; Rubin & Martin, 1994), aggressive communication (Martin, Anderson, & Thweatt, 1998), assertiveness and responsiveness, friendship cognitive flexibility, and efficacy perceptions (Martin & Anderson, 1998).

Duran and Spitzberg (1995) developed the Likert-type Cognitive Communicative Competence Scale. The items were developed to assess five factors: planning cognitions, modeling cognitions, presence cognitions, reflection cognitions, and consequence cognitions. The current version is adapted from a shorter version (Duran, Kelly, Schwager, Carone, & Stevens, 1993) and has been related to interaction involvement, communicative knowledge, and self-monitoring (Duran & Spitzberg, 1995).

Other measures are commonly considered relevant to the knowledge aspect of skilled communication, including self-monitoring (e.g., Gangestad & Snyder, 1985; Lennox & Wolfe, 1984), cognitive complexity (e.g., Burleson, 1987; Rubin & Henzl, 1984), attributional complexity (Fletcher, Danilovics, Fernandez, Peterson, & Reeder, 1986), message elaboration (Reynolds, 1997), interpersonal problem solving (e.g., Shure, 1982), empathy and role-taking ability (Spitzberg & Cupach, 1989), sensitivity to feedback (Edwards & Pledger, 1990), volitional control (Kuhl & Fuhrmann, 1998), social intelligence (e.g., Marlowe, 1985), and knowledge of the rules and possibilities of interaction (Spitzberg, 1990). Many of these measures are relatively narrow components of the knowledge component of competence.

**Self- and Other-Reference Skills Assessments**

**Communication Flexibility Scale (CFS).** Adaptability and flexibility have commonly been identified as hallmarks of competent communication skills, and yet there have been few measures developed to assess these characteristics. The CFS was developed by Martin and Rubin (1994) and consists of 14 scenarios, each of which presents a context in which experiencing new people, situations, or altering one’s communication behavior is described (e.g., “You went to a party where over 50 people attended. You have a good time, but spent most of the evening talking to one close friend rather than meeting new people”). It was related significantly to cognitive flexibility, argumentativeness, tolerance for disagreement, verbal aggressiveness, and Machiavellianism (Martin, Anderson, & Thweatt, 1998). It is also related to the noble rhetorical self but not to rhetorical sensitivity or reflection, and exhibits a small relationship with social desirability (Martin & Rubin, 1994). To some extent, social desirability motive is expected to be related to interpersonal skills. The flexibility measure taps an aspect of interpersonal skills that has received too little attention but has revealed some results contrary to what would be expected and thus needs further validation.

**Communication Functions Questionnaire (CFQ).** The CFS was developed in a program of research examining the nature of person-centered communication (Burleson, Delia, & Applegate, 1995). The items tend to be worded as other-reference descriptions that generalize across episode and setting. It consists of eight subscales, representing important functions fulfilled through communication: conflict management (e.g., “Makes me believe our relationship is strong enough to withstand any conflicts or disagreements we might have”), comforting (e.g., “Can really help..."
me work through my emotions when I’m feeling upset or depressed about something”), ego support (e.g., “Makes me feel like I’m a good person”), regulative (e.g., “Helps me see why my action broke a social rule or norm”), referential (e.g., “The capacity to express ideas in a clear, concise way”), conversational (e.g., “Can make conversations seem effortless”), narrative (e.g., “Can always get a bunch of people laughing just because he/she is so good at telling a joke or a story”), and persuasive (e.g., “Can talk me into doing things that he/she wants me to do”).

Burleson, Kunkel, Samter, and Werking (1996) found that women and men differed in their rating of the importance of six of these eight skills, but the average effect size of the difference was small. In their second study, and in research reported by Burleson, Samter, and Lucchetti (1992), friends displayed significant similarity in their communication skills, supporting a skills-mediating model of interpersonal relationship attraction. The measure is one of the few that links specific skills to functions demonstrated through other research to be relevant to interpersonal competence. The research thus far suggests that the measure has sound psychometrics and expected construct validity. As yet, however, there is little basis for presuming these skills are a comprehensive list of skills or that these skills are necessarily cast at the most useful level of abstraction (Spitzberg & Cupach, in press).

Communicative Adaptability Scale (CAS). Although somewhat a misnomer because adaptability is not directly assessed, the CAS was one of the early measures designed to encompass multiple components of competent interaction (Duran, 1983, 1992). It assesses six factors in a Likert-type scaling: social composure (e.g., “In most social situations I feel tense and constrained”), social confirmation (e.g., “I am verbally and nonverbally supportive of other people”), social experience (e.g., “I enjoy socializing with various groups of people”), appropriate disclosure (e.g., “I disclose at the same level that others disclose to me”), articulation (e.g., “I sometimes use words incorrectly”), and wit (e.g., “People think I am witty”).

Various dimensions of the CAS have been related significantly to cognitive complexity (Duran & Kelly, 1985), attractiveness (Duran & Kelly, 1988a), communication and roommate satisfaction (Duran & Zakahi, 1987, 1988), shyness (Duran & Kelly, 1989; Prisbell, 1991), communication style (Duran & Zakahi, 1984), interaction involvement (Duran & Kelly, 1988b), relationship maintenance (Prisbell, 1995), and social activity involvement (Duran & Kelly, 1992), as well as several other constructs (Duran, 1992). The dimensions of disclosure, composure, and wit were predictive of coded molecular behaviors of conversational turns, self-referencing pronouns, and other-referencing pronouns (Duran & Zakahi, 1990). The CAS has displayed consistent factor structure, acceptable psychometrics, and has generally related to other measures as expected. To date, however, these relationships have generally been with other trait measures, and therefore the predictive validity of the CAS is still in question. Furthermore, as with the Communication Functions Questionnaire, there is little rationale for the particular dimensions of the CAS. Finally, the items of the CAS tend to be cast at a fairly high level of abstraction, therefore its diagnostic value for making inferences about specific skills is limited.

Conversational Appropriateness and Effectiveness (CAE). Despite the common assumption that competent skills are appropriate and effective, few measures have attempted to assess these dimensions of behavior. The CAE was designed to measure perceptions of a particular conversational episode in terms of its appropriateness and effectiveness (Spitzberg, 1990). Appropriateness is referenced by 20 items assessing
3. METHODS OF INTERPERSONAL SKILL ASSESSMENT

the awkwardness or smoothness of behavior, embarrassment, and impressions of conversational propriety (e.g., “S/he said several things that seemed out of place in the conversation”). The effectiveness component is comprised of 20 items referencing control, goal accomplishment, and satisfaction (e.g., “I got what I wanted out of the conversation”). The initial version was semantic-differential in form (Spitzberg & Phelps, 1982) but was elaborated into a Likert-type response scale (Spitzberg & Canary, 1985).

Perceptions of appropriateness and effectiveness successfully discriminate integrative, distributive, and avoidant conflict management tactics in the perception of both self and one’s interlocutor (Canary & Spitzberg, 1987, 1989, 1990). Deep interruptions, which take speaking turns away from the person holding the floor, tend to be viewed as less appropriate and appear to have no effect on perceived effectiveness (Hawkins, 1988, 1991). The measures were highly sensitive to experimental changes in speech power (e.g., lack of tag questions, hesitations, hedges; Rose & Canary, 1988). These measures exhibited small relationships with loneliness over the last 2 weeks but were relatively unrelated to long-term loneliness (Spitzberg & Canary, 1985). The measure is virtually the only measure available to tap the dimensions most often attributed to interpersonal competence and appears to have relatively consistent factor structure. Much of the measure’s item content is relatively high in abstraction, however, thereby distancing it from the actual interpersonal skills entailed in creating the impression of appropriateness and effectiveness.

**Conversational Skills Rating Scale (CSRS).** The CSRS was developed to accommodate several assumptions about the nature of communication competence. First, if competence is a judgment of quality, then a rating scale should reflect this judgment of quality, rather than quantity. Second, if competence varies from molecular to molar, then such judgments should be separated to avoid confounding levels of inference. Third, the domain of behaviors assessed should be relatively comprehensive but also relevant to most social interactions. Fourth, the measure should be sufficiently flexible to be used in reference to self or other and in context-general as well as context-specific forms. The resulting measure (Spitzberg, 1994c, 1995) consists of five molar judgment semantic differential items (e.g., skilled–unskilled, competent–incompetent, etc.) and 25 relatively molecular items reflecting four skill clusters: altercentrism (e.g., “Speaking about partner—involvement of partner as a topic of conversation”), composure (e.g., “Shaking or nervous twitches—are not noticeable or distracting”), expressiveness (e.g., “Facial expressiveness—neither blank nor exaggerated”), and interaction management (e.g., “Speaking fluency—pauses, silences, uh, etc.”). These items are scaled on a 5-point continuum from Inadequate, Fair, Adequate, Good, to Excellent.

The factor structure has been generally supported (Spitzberg, Brookeshire, & Brunner, 1990; Spitzberg & Hurt, 1987a), although the subscales are strongly intercorrelated, thereby leading some to eschew the subscales in favor of an overall score. The measure has been significantly related to motivation, knowledge, and skills (Spitzberg & Brunner, 1991; Spitzberg & Hecht, 1984), perceptual accuracy (Powers & Spitzberg, 1986), use of humor (Graham, Papa, & Brooks, 1992), anxiety (Segrin, 1999), loneliness (Segrin, 1999; Spitzberg & Hurt, 1987b), intercultural competence (Milhouse, 1993, 1996), shyness (Prisbell, 1991), discourse strategies (Ellis, Duran, & Kelly, 1994), client outcomes in transition from welfare to work (Waldron & Lavitt, 2000), and psychosocial problems (Segrin, 1993, 1996, 1999). Ratings of self and conversational partner on the CSRS were significantly related to specific coded behaviors.
(e.g., talk duration, speaking turns, gaze) in an in vivo “waiting period” context but generally not in either a get-acquainted or role-play situation (Segrin, 1998). The CSRS is also apparently sensitive to the differing demands of these situations (Segrin, 1998).

In general, the effect sizes of the CSRS have been relatively small with trait or dispositional types of measures and moderate with more state or episodic types of behavior and measures. Its factor structure has been somewhat inconsistent across studies and may function better when the items are summed across all the items. Finally, the CSRS was designed as an episodic measure and therefore may not serve the purposes of researchers seeking to make inferences regarding a person’s trait competence.

**Interpersonal Communication Competence Scale (ICCS).** Taking a more inductive approach, Rubin and Martin (1994) developed this measure to tap the dimensions most commonly identified by the major interpersonal textbooks of the communication field. Both a long (30-item) and short (10-item) version are available, scaled on a 5-point frequency self-reference format. Both forms reflect the following skills: self-disclosure (e.g., “I allow friends to see who I really am”), empathy (e.g., “I can put myself in others’ shoes”), social relaxation (e.g., “I am comfortable in social situations”), assertiveness (e.g., “When I’ve been wronged, I confront the person who wronged me”), altercentrism (e.g., “My conversations are pretty one-sided,” reverse coded), interaction management (e.g., “My conversations are characterized by smooth shifts from one topic to the next”), expressiveness (e.g., “My friends can tell when I’m happy or sad”), supportiveness (e.g., “My communication is usually descriptive, not evaluative”), immediacy (e.g., “My friends truly believe that I care about them”), and environmental control (e.g., “I accomplish my communication goals”). The measure was strongly related to cognitive flexibility (Rubin & Martin, 1994) and moderately related to dyadic roommate satisfaction (Martin & Anderson, 1995). The ICCS has the intuitive appeal of strong face validity and representativeness. There is little basis however, to conclude that textbooks in a discipline reflect the skills identified by the more scholarly literature as actually related to competence. Furthermore, there are significant variations in item abstraction. Finally, the measure has simply not been studied extensively enough to determine its construct validity.

**Interpersonal Competence Questionnaire (ICQ).** Developed by Buhrmester, Furman, Wittenberg, and Reis (1988), this is a dispositional measure intended to assess competence in five domains: initiation (e.g., “Introducing yourself to someone you might like to get to know”), self-disclosure (e.g., “Revealing something private about yourself when talking to a close friend about personal matters”), negative assertion (e.g., “Asking someone you’ve been dating to change an irritating mannerism”), advice and guidance (e.g., “Not exploding at a close companion—even when it is justified—to avoid a damaging conflict”), and conflict resolution (e.g., “Helping a close companion cope with family or roommate problems”). The items are scaled on a 5-point scale from “I’m poor at this” to “I’m extremely good at this.” Its vocabulary has been adapted for adolescent populations (Buhrmester, 1990).

The factor structure of the ICQ appears resilient to adaptation to other relationships and populations (e.g., Theriault, 1997). It is significantly related to college adjustment (Shaver, Furman, & Buhrmester, 1985), popularity, dating initiation and frequency, depression, well-being, loneliness, dating skill, assertiveness, social anxiety, shyness, masculine ideology, physical attractiveness, and emotional expressiveness (Bruch,
Berko, & Haase, 1998; Buhrmester et al., 1988) and to various measures of motivation, knowledge, and interpersonal skills (Spitzberg, 1990), with effect sizes ranging from small to moderate and varying considerably across dimensions. Although the subscales produced somewhat inconsistent results and varied by respondents’ gender and by gender composition of the dyad, the ICQ was generally negatively associated with emotional reactivity, as expected (Bartle-Haring & Sabatelli, 1997). The measure has shown significant relationships to stress over time and small predictive relationship to subsequent psychopathology (Herzberg et al., 1998). Its adolescent self-report form showed consistently small to large relationships with sociability, hostility, anxiety, depression, and self-esteem, although the preadolescent form and “friend-rated” form of the measure showed more inconsistent results (Buhrmester, 1990).

The ICQ is one of the few measures developed to tap “relational competence.” As such, it represents a fairly narrow domain of contexts and skills. The relative importance of the skills it taps shift considerably from one study to another, making generalizations about its subscale structure difficult. Finally, its item content is cast at a fairly high level of abstraction, meaning that it provides relatively little diagnostic information about actual behaviors involved in these interpersonal contexts. Within the domain of relational competence, however, it has received relatively strong support as a trait measure of interpersonal skills.

Social Performance Survey Schedule (SPSS). The SPSS is a 100-item Likert-scaled measure intended to assess positive (50 items; e.g., “Has eye contact when speaking”) and negative (50 items; e.g., “Puts himself/herself down”) interaction behaviors (Lowe, 1982, 1985; Lowe & Cautela, 1978). Factor analysis has exhibited a multidimensional structure (Lowe & D’Ilio, 1985), but this structure has not been replicated or consistently employed in research. An abbreviated version of the measure has shown statistically significant discriminant validity (Fingeret, Monti, & Paxson, 1985; Wessberg et al., 1981) and convergence across observers and contexts (Monti, 1983; Wessberg et al., 1981). The measure has also successfully predicted several conversational behaviors (Miller & Funabiki, 1983). Student leaders perceive themselves as engaging in significantly higher rates of positive behaviors than their parents perceived of their children, and girls report higher frequency of positive, and lower frequency of negative, behaviors than boys (D’Ilio & Karnes, 1987, 1992; Karnes & D’Ilio, 1989). The positive subscale was strongly correlated with depression, social activity, observer ratings of social skill, and introversion (Lowe, 1982). The measure exhibited convergent validity in predicting number of interactions, number of friends, peer likeability, talk time, eye contact, and self and observer in vivo skill ratings (Lowe, 1985). This scale, however, was also strongly related to social desirability (Lowe, 1982). The SPSS has withstood more rigorous tests than most competence measures. Its limitations are that it is long and relatively undifferentiated.

Social Skills Inventory (SSI). One of the few assessments with an explicitly theoretical approach, Riggio (1986) reported a dispositional self-reference measure intended to represent three dimensions of skills. Skills are conceptualized as serving expressivity (i.e., sending), sensitivity (i.e., receiving), or control (i.e., monitoring) functions. These skills involve sending, receiving, or monitoring emotional messages (i.e., communicating messages of affect, attitude, and status relationship) and social messages (i.e., verbal fluency and facility). The measure consists of 105 items scaled on a standard 5-point Likert-type response continuum, although some studies exclude the control scales.
The measure has generally produced expected convergent and discriminant coefficients with personality measures, attractiveness, social anxiety, self-consciousness, and nonverbal skills (Riggio, 1986). The measure also correlated with employment experience, social experience, and shyness (Riggio, 1986). The SSI was correlated to self-esteem, anxiety, and loneliness but not to locus of control or well-being (Riggio, Throckmorton, & DePaola, 1990; Vandeputte et al., 1999). The measure correlated largely as expected with various measures of communicative motivation, knowledge, and skill (Spitzberg, 1990). The measure exhibited few significant relationships between parents’ and children’s social skills (Segrin, 1994). The measure has displayed small but significant predictive and moderating effects in predicting psychosocial and academic problems over time (Segrin & Flora, 2000). There appear to be no birth order effects on the SSI scales (Riggio & Sotoodeh, 1989). Subscale composites produced moderate to large relationships with measures of empathy (Riggio, Tucker, & Coffaro, 1989) and small to moderate relations with use of others for coping types of social support (Riggio & Zimmerman, 1991).

Certain subscales and the total measure have predicted believability in deception tasks and deception skill (Burgoon, Buller, & Guerrero, 1995; Burgoon, Buller, Guerrero, & Feldman, 1994; Riggio, Tucker, & Throckmorton, 1988). Some of its scales display significant correlations with social desirability (Riggio, 1986). The SSI has received extensive application and performed very well. It has a clearly defined subscale structure and demonstrates solid psychometrics and validity coefficients. To date different applications have applied it in various ways, however, excluding subscales or summing across subscales, making generalizations about the validity of specific subscales of the SSI problematic. Finally, the contents of items vary considerably in their abstraction both within and across subscales, meaning that very different levels of inference are invoked.

**Miscellaneous Assessments.** A few additional measures deserve brief mention. These are scales that either have not received sufficient attention to review in full or that assess skills in contexts that are narrowly related to competent interaction. Schrader’s (1990; Schrader & Liska, 1991) Refined Measure of Interpersonal Communication Competence is a 39-item measure based on items that best discriminated competent versus incompetent communicators. The Communicator Competence Questionnaire was developed to assess encoding (e.g., “My subordinate has a good command of language”) and decoding (e.g., “My subordinate is a good listener”) dimensions in the organizational context (Monge, Bachman, Dillard, & Eisenberg, 1982). Lorr, Youniss, and Stefic (1991; Schill, 1995) developed a 128-item, true–false Social Relations Survey of social skills assessing eight hypothesized bipolar domains: social assertiveness, directiveness, defense of rights, confidence, perceived approval, expression of positive feelings, social approval need, and empathy. Gambrill (1995) reported a 24-item Social Competence Scale referencing a variety of social behaviors (e.g., “offering friendly reactions,” “maintaining conversations”), scaled in a straightforward competence continuum (i.e., not at all competent to very competent). For somewhat molar level assessment of episode-specific interpersonal competence, the Self-Rated Competence and Rating of Alter-Competence measures have been widely used (see Meeks, Hendrick, & Hendrick, 1998; Perotti & DeWine, 1987; Segrin, 1994; Spitzberg, 1988; Spitzberg & Cupach, 1989). Finally, although the measure is in its infancy, the development steps involved are so comprehensive that the Interpersonal Communication Competence Inventory should be mentioned (Bubaš, 2000; Bubaš, Bratko, & Marusic, 1999). It was developed with a relatively comprehensive attention to previous work in competence, and exhibits expected validity coefficients
with the Communication Functions Questionnaire (Burleson & Samter, 1990), the Interpersonal Competence Questionnaire (Buhrmester et al., 1988), the Social Skills Inventory (Riggio, 1986), and an Interpersonal Communication Skills Questionnaire previously developed by Bubas. The ICCI reveals two second-order factors (communication effectiveness, other-orientedness) and up to six specific factors (decoding and encoding, social relaxation, expressivity, intentionality).

Several measures of interpersonal skills have also been developed to reference particular contexts. Inderbitzen and Foster (1992) developed a Teenage Inventory of Social Skills. In the dating context, measures such as the Dating and Assertion Questionnaire (Levenson & Gottman, 1978), Heterosexual Assessment Inventory for Women (Kolko, 1985), and Survey of Heterosexual Interactions (Williams & Ciminero, 1978) have been used extensively. Several scales have been reported that assess intercultural communication competence (e.g., Martin & Hammer, 1989). In the counseling context, Remer (1978) reports a Potential Interpersonal Competence Scale. In the marital context, the Communication Patterns Questionnaire (Heavey, Larson, Zumtobel, & Christensen, 1996; Noller & White, 1990) has successfully predicted conflict-relevant marital outcomes. Similarly, the argumentativeness–aggressiveness construct has been applied as a proxy for communicative competence in conflict situations (e.g., Onyekwere, Rubin, & Infante, 1991).

Direct Skills Assessments

Role-Play and Scenario Methods. Role-play methods represent a host of techniques for eliciting sample response behaviors from subjects. Role-play methods are not technically assessment instruments; however, so much behavior therapy and counseling research yokes similar types of assessment scales to role-play scenarios, and role-play scenarios are crafted carefully so as to elicit certain types of skills (e.g., assertiveness, heterosocial, etc.), that role-play methods have become widely considered a form of assessment. The behaviors elicited from role-play stimuli are subsequently either coded for the occurrence of behaviors assumed relevant to competence (e.g., amount of eye contact, number of gestures, etc.) or evaluated in terms of competence (e.g., unskilled–skilled, unattractive–attractive). The scenarios developed to elicit such response behaviors are developed to be reasonably relevant, realistic, representative, and engaging for the respondents.

Role-play methods have been the subject of considerable research examining their validity (e.g., Ammerman & Hersen, 1986; Bellack et al., 1978, 1979; Frisch & Higgins, 1986; Kazdin, Matson, & Esveldt-Dawson, 1984; Kern, 1982; Kern et al., 1983; Kolotkin & Wielkiewicz, 1984; Letherman et al., 1984; Letherman, Williamson, Moody, & Wozniak, 1986; Mahaney & Kern, 1983; McNamara & Blumer, 1982; St. Lawrence, Hughes, et al., 1983; St. Lawrence, Kirksey, et al., 1983). Most research has shown that role-play methods are highly sensitive to various forms of demand effects (Spitzberg, 1991), including degree of standardization of situational stimuli (Chiauzzi, Heimberg, Becker, & Gansler, 1985), expectation set of instructions (Ammerman & Hersen, 1986; Spitzberg & Chandler, 1987), mode of stimulus presentation (Galassi & Galassi, 1976; Perlmutter, Paddock, & Duke, 1985; Remer, 1978; Smit & van der Molen, 1996), confederate prompt delivery style (Mahaney & Kern, 1983; Steinberg, Curran, Bell, Paxson, & Munroe, 1982), gender of confederate (Eisler, Hersen, Miller, & Blanchard, 1975), sex of ratee (Gormally, 1982), familiarity with ratee (Gormally, 1982), and race of ratee (Hrop & Rakos, 1985; Turner, Beidel, Hersen, & Bellack, 1984). Nevertheless, carefully developed role-play methods can provide a relatively standardized approach to eliciting observable social behavior.
from respondents for evaluating social competence. Some of the more relevant and widely used role-play methods are reviewed below.

**Simulated Social Interaction Test (SSIT).** The SSIT consists of eight role-play situation descriptions (Curran, 1982). The eight situations represent a range of potentially problematic social encounters: disapproval or criticism, social assertiveness or visibility, confrontation and anger expression, heterosexual conflict, interpersonal warmth, conflict with or rejection by parent or relative, interpersonal loss, and receiving compliments. Examples follow:

*Narrator:* You are at work, and one of your bosses has just finished inspecting one of the jobs that you have completed. He says to you,

*Confederate:* “That’s a pretty sloppy job. I think you could have done better.”

*Narrator:* You are at a party, and you notice a woman has been watching you all evening. Later, she walks up to you and says,

*Confederate:* “Hi, my name is Jean.”

These prompts can be played on audiotape, videotape, or presented by an assessor or by confederate. The subject’s responses are typically recorded, and the responses to the eight situations are rated on two 11-point scales assessing social anxiety and social skill and summed across situations.

Research has examined the SSIT among a variety of subject populations, and it has been compared across various types of rater expertise (Curran, Monti, et al., 1980; Farrell, Curran, Zwick, & Monti, 1984; Fingeret, Monti, & Paxson, 1983; Fingeret et al., 1985; Mersch, Breukers, & Emmelkamp, 1992; Monti, 1983; Monti et al., 1984; Monti, Curran, Corriveau, DeLancy, & Hagerman, 1980; Monti & Fingeret, 1987; Monti, Wallander, Ahern, Abrams, & Munroe, 1983; Monti, Zwick, & Warzak, 1986; Steinberg et al., 1982; Wallander, Curran, & Myers, 1983; Wessberg et al., 1981). The research collectively indicates that the SSIT is a sensitive and valid measure of social skills.

**Social Interaction Test.** One of the persistent problems in assessing competence is that mere output of behaviors is rarely a measure of competence, because too little or too much of virtually any behavior can be incompetent. Trower, Bryant, and Argyle (1978) attempted to cope with this by developing an elaborate rating scale that encompasses parallel descriptions of too much and too little. The scale is applied to behavior stimulated in an in vivo “tell us about yourself” task in which confederate behaviors are manipulated. The rating scale has been adapted to alternative applications, and its item content adapted, since its original incarnation. A sample item, regarding volume of speech, is as follows:

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Normal volume</td>
</tr>
<tr>
<td>1(a)</td>
<td>Quiet but can be heard without difficulty</td>
</tr>
<tr>
<td>1(b)</td>
<td>Rather loud but not unpleasant</td>
</tr>
<tr>
<td>2(a)</td>
<td>Too quiet and difficult to hear</td>
</tr>
<tr>
<td>2(b)</td>
<td>Too loud and rather unpleasant</td>
</tr>
<tr>
<td>3(a)</td>
<td>Abnormally quiet and often inaudible</td>
</tr>
<tr>
<td>3(b)</td>
<td>Abnormally loud and unpleasant</td>
</tr>
<tr>
<td>4(a)</td>
<td>Inaudible</td>
</tr>
<tr>
<td>4(b)</td>
<td>Extremely loud (shouting).</td>
</tr>
</tbody>
</table>
Adapted forms of the measure have been used in a variety of applications (e.g., Caballo & Buela, 1988; Turner, Beidel, Dancu, & Stanley, 1989). The Fydrich et al. (1998) adaptation, for example, reduced the 29 item sets into five: gaze, vocal quality, length, discomfort, and conversation flow. The rating scale has been employed to examine self versus other impressions of competence (Spitzberg & Cupach, 1985), molar versus molecular impressions of skills (Caballo & Buela, 1988), and social anxiety (Turner et al., 1984). Because the rating scale and original stimulus context have rarely been applied consistently, it is difficult to generalize about the measure’s validity, but its conceptual rigor and comprehensiveness show considerable promise.

**Marital Communication Assessments.** There is an extensive literature on assessing communication skills in the marital or intimate partner context. Many of these systems involve the rating or coding of marital interactions (Gottman & Notarius, 2000). These interactions generally surround problem-solving tasks but can be more mundane or positive in socioemotional content. Some examples of direct interpersonal skills assessments in the couples’ context include the Clinician Rating of Adult Communication (Basco, Birchler, Kalal, Talbott, & Slater, 1991), the Marital Interaction Coding System (Heyman, Weiss, & Eddy, 1995), the Couples Interaction Scoring System (Gottman, 1979; Gottman & Rushe, 1995), and the Specific Affect coding system (Waltz, Babcock, Jacobsen, & Gottman, 2000). Other observational measures have been developed for rating or coding expressed affect in marriages (e.g., Gottman, 1994; Krokoff, 1991; Smith, Vivian, & Leary, 1990). These systems typically identify certain types of positive and negative affective exchanges, as well as interactional moves that work away from mutually respectful task-oriented problem solving (Butler & Wampler, 1999). Finally, the Communication Box test has been employed successfully to assess predictive accuracy, perceptual accuracy, and communication effectiveness in interacting couples (e.g., Burleson & Denton, 1992, 1997; Denton, Burleson, & Sprenkle, 1995; Gottman, 1994). Interactants take a turn at talk and then note what they intended the effect of that message to be on the co-interactant and what effect they predicted the message would actually have. The co-interactant then indicates what effect the message had and then engages in his or her own turn at talk. Measures are formulated from the discrepancies between the partner intended, predicted, and actual meanings.

**Children’s Sociometric and Peer Assessments.** Some populations are not trusted to provide self-assessments of their social skills due to the particular nature of their inability to perceive the world in normative ways (e.g., people with schizophrenia). Other populations are included in the assessment process but are rarely trusted with the sole assessment role. One such population is children, for whom issues of popularity among peers is a highly relevant marker of interpersonal competence but is unlikely to be validly represented by self-assessment alone. There is an extensive literature on children’s social and interpersonal skills (for empirically based reviews, see, e.g., Caldarella & Merrell, 1997; Newcomb, Bukowski, & Pattee, 1993; also see Samter, this volume). Numerous studies of children’s social skills have employed some variation of peer assessment, which most typically takes the form of sociometric rankings of popularity, attractiveness, likability, and so forth (Inderbitzen, 1994). These are viewed here as direct methods because the evaluator is presumed to be referencing experience with direct and observed interaction with the people being ranked. Although such approaches rarely refer to specific skills, they are often used as a measure of children’s basic interpersonal skills. The typical approach in such a study
is to locate a relatively closed social system (e.g., a classroom or fraternity) and ask subjects to rank order their peers in this system according to some dimension of competence (e.g., social attractiveness). The average rank a subject receives represents that person’s social competence. These measures are often compared with or combined with ratings of specific social skills observed by teachers, school personnel, parents, or peers (e.g., Cairns, Leung, Gest, & Cairns, 1995; Demaray et al., 1995; Feng & Cartledge, 1996; Flanagan et al., 1996; Matson et al., 1995; Newcomb et al., 1993). Other approaches have formulated coding systems (Santoyo, 1996) or elaborate rating systems (e.g., Bain, 1991) to assess interaction skills (e.g., Santoyo, 1996), but such methods have not been widely adopted.

Miscellaneous Assessments. Some direct assessments deserve brief note because of their relevance or relative lack of attention. Farrell and colleagues developed a set of “intermediate-level” rating scales that anchored various molecular behaviors on scales of appropriateness, along with descriptors and behavioral exemplars (Farrell et al., 1984, 1985). Haley (1985) developed 28 role-play scenes to assess social skills in negative assertion, positive assertion, and initiating social contact situations. Many efforts have been made to employ a “standardized patient” role-play assessment of interaction and patient-interviewing skills of current and prospective physicians (e.g., Boulet et al., 1998; Cohen, Colliver, Marcy, Fried, & Swartz, 1996; Gruppen et al., 1997). In the standardized patient assessment, subjects are asked to respond to a series of hypothetical patient presentations, which can be presented in live, recorded, or written forms, and the responses to these scenarios are then evaluated according to third-party skill ratings. Ralph and colleagues (Ralph, 1990; Ralph & Lee, 1994; Ralph & Thorne, 1993) have applied a verbal interaction coding system to assess competence in initiating and maintaining topics in interview situations.

Alternative Assessments

There are alternative approaches to assessment that are not easily categorized by the Behavioral Assessment Grid. For example, there is extensive discussion of portfolio assessment of communication skills in instructional contexts, and such approaches have relevance to interpersonal skills (e.g., Jacobson, Sleicher, & Maureen, 1999). Efforts continue to develop an automated or computer-based program for social skills assessment (Holsbrink-Engels, 1997; Muehlenhard & McFall, 1983; O’Neil, Allred, & Baker, 1997). At this point, these efforts tend to be too diverse and nascent to review here. Nevertheless, they also suggest that there are still unexplored horizons in the assessment of interpersonal skills. Success in pursuing these horizons may well depend on the resolution of a number of key issues that still present significant obstacles to the valid assessment of interpersonal skills. Some of these issues are reviewed next.

KEY ISSUES IN INTERPERSONAL SKILL ASSESSMENT

This highly selective review illustrates how diverse and extensive the options are for assessing social skills. In the context of so many options, the crucial question arises as to which assessment is “best.” Selection of any assessment for any given project depends significantly on how one answers six key questions, which are discussed here under the abstracted rubric of what, when, where, who, how, and why (Spitzberg, 1987).
What: Which Competence Domain(s) Will Be Assessed?

This basic question receives surprisingly little detailed consideration in assessment projects. Spitzberg and Cupach (1989; Spitzberg, 1994b) identified well over 100 factor-analytically derived labels of skills or dimensions attributed to interpersonal competence or skills. Does interpersonal competence consist of assertiveness, self-disclosure, heterosocial initiation, composure, other-orientation, empathy, role-taking, sensitivity, listening, attentiveness, articulation, wit, responsiveness, creativity, adaptability, control, expressiveness, clarity, understanding, or some other combination of social behaviors? Furthermore, even these skills vary considerably in terms of the more molecular skills that comprise them. Spitzberg and Cupach (in press) have suggested a perspective that categorizes molecular skills (e.g., asking questions, laughter) according to several more intermediate level skills (e.g., attentiveness, expressiveness), which then relate in multiple ways to higher order functional skills (e.g., empathy, excitement), which relate to even higher order functions (e.g., intimacy, novelty), and finally to the most fundamental interpersonal functions (i.e., moving toward or with another, moving away from another, moving against another). Although still highly speculative, this approach at least recognizes the importance of attempting to map a place for the entire terrain of potential skills at various levels of abstraction.

Most assessment projects would ordinarily employ theory to guide selection of skills. To date, however, there still are no widely accepted theoretical models that specify what skills comprise the essential competencies of social interaction. Furthermore, there are almost no models that specify the interrelationships among the components of social skill. Only a few studies have attempted to develop truly multivariate models of social skills in the larger social process (Bruch et al., 1998; Canary & Spitzberg, 1989; Lamke, Sollie, Durbin, & Fitzpatrick, 1994; Rubin, Martin, Bruning, & Powers, 1993). Even these models, however, tend to view social skills themselves as univariate constructs. For example, if competence is comprised of both empathy and assertiveness, how can such seemingly incompatible skills be combined in a given context (Lowe & Storm, 1986; Spitzberg et al., 1994)? Such complex multidimensional interrelationships need to be discerned both conceptually and empirically.

Why: What Will the Relation of Assessment Be to Valid Social Outcomes?

The ultimate question for assessment projects is: Assessment for what? Why are people being assessed in the first place, and what uses can be made of the assessment? The answers to these questions are not as obvious as they might seem. In educational and clinical contexts, the general assumption is that assessment is being conducted to guide opportunities for remediation and enrichment of those being assessed. But in the educational context, assessment also can be used for accountability, as in evaluation or action research. In the clinical context, assessment can serve the purposes of publication, prestige, and grinding personal theoretical axes. Assessments developed for basic research are often ignored by those seeking more applied measurement schemes, and vice versa. Increasingly, scholars and practitioners are calling for more social validation of assessments, which requires establishing ecological and representational validity for those being assessed. Much research, for example, has presupposed the relevance of assertiveness to people’s lives without ever asking people to identify aspects of their everyday lives in which more competent assertiveness
could have made a substantive difference in their social outcomes. It may be that assessments too often reflect what is relevant to the researcher, clinician, or educator, rather than what is relevant and important to the interactant.

**When: Will Assessments Be Dispositional (Trait) or Episodic (State)?**

Is competence a state or a trait? Are interpersonal skills something that are manifest across time and contexts, or are skills contextually specific and unique to a given episode of interaction? It is generally considered axiomatic that competence is contextual (Spitzberg & Brunner, 1991). However, one of the most common purposes of assessment is to produce a diagnostic estimate of a person’s abilities across contexts. If, in contrast, competence or skill is entirely episodic, then any assessment can be considered valid only in the context itself. To some extent this is an empirical question, and to date the evidence is not very encouraging (Segrin, 1998; Spitzberg, 1987, 1991). Dispositional and traitlike measures generally have not produced strong correlations with behavioral performance. Furthermore, episodic assessments such as brief role-play methods often do not generalize to other in vivo or naturalistic situations (Bellack et al., 1978, 1979). Although generally overlooked in the initial stages of assessment development, it is likely to matter what time frame the subject is asked to consider when rating competence. When assessing one’s own competence, it makes a difference whether the last 2 weeks, 2 years, or 2 decades represent the time frame for assessment. Communication skills are developmental and contextually sensitive. Therefore, a person’s competence during the college years may be different from his or her high school years, and even the freshman year may be different from the senior year (Rubin & Graham, 1988; Rubin, Graham, & Mignerey, 1990). Competence evaluations may become more self-focused, positive, and molar over time (i.e., specific behaviors are forgotten, whereas general themes and evaluations are recalled) over time (Spitzberg, 1987). Therefore, any approach to assessment must address the issue of the time frame in which competence is assessed.

**Where: Will Skills and Competence Be Framed Within a Specified Context?**

The axiomatic assumption that competence is contextual refers to both time and “place.” These are related but distinct issues. Contexts can be conceptualized along cultural, chronemic, relational, situational, or functional dimensions or typologies (Spitzberg, 2000; Spitzberg & Brunner, 1991). Research makes clear that competence evaluation does indeed vary along these dimensions and types (Argyle et al., 1981; Burgoon, Stern, & Dillman, 1995; Forgas, 1979; Heise, 1979). It follows, therefore, that any assessment project must determine what features of the context are relevant to the display of competence and design those features into the assessment method.

**Who: Will Assessment of Competence and Skills Be Made by Interactants, Third-Parties, or Both?**

Some populations are assumed to be limited in their ability to provide self-referential data on competence. The mentally disabled, mentally ill, and infants, for example, generally are viewed as lacking the self-reflective ability to comment validly on their own social competence. Other populations are simply viewed as systematically biased in their self-assessments (e.g., depressed, lonely, etc.). “Bias” in evaluation of
competence can be viewed as intrinsic to any vantage point, however. Research is clear, for example, that self perceives competence differently from one’s interactional partner (e.g., Spitzberg & Cupach, 1985; Spitzberg & Hurt, 1987a; Sypher & Sypher, 1984) and from third-party observers or subsequent raters (Farrell, Mariotto, Cooper, Curran, & Wallander, 1979; Nelson, Hayes, Felton, & Jarrett, 1985). Furthermore, such disparities are exacerbated by such interactant states as depression, loneliness, and anxiety (e.g., Segrin, 1998) and rater characteristics such as gender (Conger et al., 1980; Spitzberg & Hurt, 1987a) and ethnicity (Turner et al., 1984). Rater training may simply be a way of substituting one set of biases for another (Bernardin & Pence, 1980; Conger, Wallander, Ward, & Farrell, 1980; Corriveau et al., 1981). Finally, people are likely to assess the competence of a well-known person differently from a stranger, and various relational characteristics are likely to affect such evaluations (e.g., status difference, intimacy, affect, etc.). Thus, any attempt at developing a sound assessment presupposes that a reasoned decision has been reached regarding who is an appropriate source for competence assessment. When the skills being assessed involve episodes of interaction in a person’s past, self-report, or other-report by someone well known to the subject, is likely to be most useful. Other-report and direct observation methods are more likely to be useful when the skills being assessed are very context specific and thus capable of being displayed in a single episode of interaction. Such methods will also be preferred when the subject being assessed is unable to provide an accurate appraisal, such as with small children or with disordered adults. Role-play methods will be useful when the skills assessed and the contexts that elicit them can be well defined.

How: How Will Assessment Be Operationalized?

Many practical issues could be examined under this heading (e.g., how are instructions written, whether written scenarios are equivalent to audiotaped scenarios, etc.). Two issues are discussed here because they have received the least attention from assessment research: quantity versus quality of scaling and curvilinearity of scaling.

One of the first questions to answer is whether to scale items or skills in terms of their frequency or duration of occurrence or in terms of their quality of performance. Many role-play methods, for example, code the occurrence of certain behaviors (e.g., amount of eye contact, number of questions asked, etc.) and then correlate these with molar evaluations of the subject’s overall episodic performance. This begs the question, however, as to what is competence in these studies. If it is the evaluations, then any unexplained variance in the ratings is unaccounted for competence. It exists, but its basis is unknown. Furthermore, if competence is in the molar evaluations, then it is an inference of those particular third parties and raises some of the “who” issues discussed above. If, in contrast, it exists in the behaviors or skills themselves, then what is the role of the molar evaluations? Eventually, assessment will need to “be directed at transforming the results from one frame of reference to another by developing mappings from the more objective measurement frames to those represented by participant and observer judgments” (Cappella, 1991, p. 111).

A more troublesome aspect of scaling issues is curvilinearity (Spitzberg, 1987). Virtually any behavior or skill is incompetent at the extremes of disuse or excessive use (Wiemann, 1977). Both a conversation with no eye contact and a conversation with 100% eye contact are likely to be viewed as incompetent by most partners or third parties. Yet scaling rarely reflects the possibility of this curvilinearity. A few measures have built in the possibility of curvilinearity (e.g., Trower et al., 1978), and
other scaling systems have overcome it by making the continuum of evaluation one of quality rather than quantity (e.g., Spitzberg, 1995). Regardless, if the quantity with which skills are displayed is viewed as relevant to the assessment method, then some manner of coping with the possibilities of curvilinearity are necessary.

A second issue related to “how” assessment is operationalized concerns the notion of competence thresholds. Competence can be arrayed along a continuum from “minimal” to an “optimal” achievement. As such, any assessment approach must address what ultimate criterion of competence must be displayed. Using the example of the “eye test,” an assessment will reflect whether vision is assessed under optimal conditions, whether vision is instead challenged to determine its adequacy under less than optimum conditions, or whether only minimum levels of vision need be demonstrated. Similarly, the behavior-analytic approach to competence tends to reflect the assumption that competence is best assessed under conditions in which people’s responses are substantially taxed by the complexity and difficulty of the situation. In contrast, many assessments merely ask people to judge their own ability to “make friends” or “make eye contact.” Such measures may be interpreted at relatively low levels of expectation for performance. That is, if the person does not lose friends every day, it may seem that he or she “makes friends” competently. However, not losing friends frequently is quite different from establishing and maintaining highly satisfying friendships whenever needed and under awkward or trying circumstances (e.g., moving away to college, at a club, etc.). Assessment approaches need to address the conceptual implications of minimal versus optimal competence and how such implications alter the types of scaling used and the ways in which stimuli and scenarios are developed.

CONCLUSION

The various assessment maps of interpersonal skills and competence are complex. But the territory to which the maps refer is inevitably more complex. The terrain of social interaction skills is clearly too complex for any one map to permit complete navigation of its subtleties. Attempts to use the landmarks on the map as representations of what is “really there” are risky at best and seriously misdirected at worst. An example from the past 20 years provides a good illustration. In the socially liberating zeitgeist of the 1960s in the United States armed with the motor skills model of social skills and the general pathos of emotional expressiveness, the assertiveness movement was embraced by clinical psychology (Galassi, Galassi, & Vedder, 1981). A more open, honest, and reasoned basis for societal interaction was envisioned. Elaborate training and assessment programs were instituted, and the behavior therapy literatures burgeoned with assertiveness coding systems, rating scales, self-report measures, and role-play scenarios.

In particular, a methodological paradigm arose befitting the law of the hammer. The paradigm study involved subjects in assertiveness role-play scenarios. The subjects' behavioral responses to these scenarios would be videotaped, coded for objective behaviors (e.g., eye contact, talk time, etc.), and rated by other third parties for overall social skillfulness. The molecular coded behaviors would then be correlated with the molar evaluations of skillfulness. Study after study confirmed that the more assertive one’s verbal and nonverbal behavior, the more socially skilled the interactant was perceived to be by uninvolved third parties. However, no one seemed to be asking how the interactant’s assertiveness was perceived by the conversational partners,
confederates, or significant others (e.g., spouses, friends, etc.). When programs of research began investigating just these types of questions, it was found that interactants whose verbal and nonverbal skills were more assertive were perceived as more effective, but also often as less attractive, likable, and appropriate (Spitzberg et al., 1994). In essence, when the underpinnings of the assessment paradigm are changed, a Dale Carnegie nightmare emerged in which large numbers of people were being trained “how to influence people and lose friends.” Assessment is inherently ideological and theoretical, and before any assessment project is undertaken, the developers and users would do well to examine these underpinnings.

REFERENCES


3. METHODS OF INTERPERSONAL SKILL ASSESSMENT


Sheridan, T. (1762). *A course of lectures on elocution: Together with two dissertations on language; and some other tracts relative to those subjects*. London: W. Strahan.


3. METHODS OF INTERPERSONAL SKILL ASSESSMENT


