The Riverside Behavioral Q-sort: A Tool for the Description of Social Behavior

David C. Funder and R. Michael Furr
University of California, Riverside
C. Randall Colvin
Northeastern University

ABSTRACT The Riverside Behavioral Q-sort (RBQ) is a flexible technique for gathering a wide-ranging description of the behavior of individuals in dyadic social interaction. Ratings of RBQ items can attain adequate reliability to reflect behavioral effects of experimental manipulations and to manifest meaningful correlations with a variety of personality characteristics. The RBQ's flexibility, validity, and relative ease of use may facilitate the more frequent inclusion of behavioral data in personality and social psychology.

"Psychology can be defined as the *scientific study of behavior and mental processes*" (Atkinson, Atkinson, Smith, & Bem, 1993, p. 4).

"Psychology is formally defined as the scientific study of the behavior of individuals and their mental processes" (Zimbardo & Weber, 1994, p. 5). Behavior is of central importance in many conceptualizations of psychology, as illustrated by the way introductory textbooks commonly incorporate "the study of behavior" into the very definition of the field.

A large number of individuals participated in the development of the Riverside Behavioral Q-sort, including Melinda Blackman, David Kolar, Daniel Ozer, Carl Sneed, Jana Spain, Mary Verdier, and many other past and present students and colleagues. This research was supported by grant R01-MH42427 from the National Institute of Mental Health. Correspondence concerning this article may be addressed to David C. Funder, Department of Psychology, University of California, Riverside, CA 92521. E-mail may be sent to funder@citrus.ucr.edu.

Journal of Personality 68:3, June 2000.

Moreover, although psychology is widely acknowledged to encompass both behavior and mental processes, the only empirical window into mental processes is—barring ESP—through the observation of behavior of some sort. This behavior may include verbal reports (including questionnaire responses), nonverbal indicators such as response latencies or body movements, or overt social behaviors.

The data obtained by research in personality and social psychology frequently include subjects' self-descriptions, reports of perceptions and memories, and responses to questionnaires. For example, participants may be asked to describe their general behavioral patterns, to relate their opinions of themselves, to estimate the frequency with which they emit certain behaviors, or to predict what they would do in certain situations. While data like these are valuable and have been profitably used in a variety of domains, they all rest upon self-report.

Self-report is an indispensable part of the research arsenal, but has some obvious limitations. For example, subjects may lack self-knowledge, be unable to predict what they would do in unfamiliar settings, distort their self-image to maintain self-esteem, or simply be unwilling to share certain secrets about themselves with psychological researchers. For these and other reasons, psychological research must at least occasionally reach beyond what subjects say, and attempt to assess what they actually do.

Commonly employed measures of behavior that go beyond self-report include response latencies (e.g., how long it takes a subject to respond to a verbal cue or how long he or she is willing to wait for a delayed reward), imposed choices (e.g., which of two rooms a subject prefers to wait in), and other single reactions (e.g., how much shock a subject administers to a confederate of the experimenter). Measures such as these are valuable and the studies that employ them have provided important insights into personality and social processes. They typically are limited, however, in two ways. First, some of the behaviors measured, while informative about theoretical models of response, may be intrinsically uninteresting. For example, a response latency measured in milliseconds may be informative about social information processing, but is not important in and of itself. Second and perhaps more important, the modal number of behaviors measured in studies that include behavioral measures at all, is one. In the typical case, a single behavioral indicator of a hypothesized underlying process is observed and recorded, and everything else the subject might be doing at the same time is ignored.

The window on behavior provided in such a study may be useful, but it is extremely narrow.

The narrowness of the usual empirical view on social behavior has occasionally left personality psychology in a vulnerable position. When Mischel (1968) challenged the field to provide evidence that self-report measures of personality traits were correlated with *behavior*, not just other self-report measures, embarrassingly little data were available on which to build a response (Block, 1977). While that controversy eventually dissipated (Kenrick & Funder, 1988), personality psychology has remained slow to build its inventory of demonstrated associations between important aspects of personality, on the one hand, and overt social behaviors, on the other.

The neighboring field of social psychology has been almost as slow to include comprehensive assessments of behavior into its research. One part of social psychology, the study of "social cognition," has moved almost entirely away from the overt social behaviors that once were the raison d'être of the field. Remaining research in social psychology has stuck rather closely to the time-honored strategy of measuring but a single behavioral dependent variable in each study. The result is a field that has learned much about the situational independent variables that may affect behavior, but rather less about the actual range of behaviors that these variables influence.

The reasons for this state of affairs are not difficult to discern. First, the study of questionnaire responses, subjective reports of memories and perceptions, and single overt behaviors has been sufficient to bring psychology a long way. Much can be discovered using these methods, so the need to go beyond them may sometimes be viewed as less than urgent by busy, resource-strapped investigators. Second, few techniques for the comprehensive measurement of overt behavior have been developed, and some of those that do exist are burdensome (requiring, for example, extensive training of behavioral coders and hours of coding for each minute of behavior). More discouraging, sometimes these extraordinarily expensive techniques have seemed to yield little more substantive knowledge about psychology than that which can be obtained using less comprehensive methods.

The purpose of the present article is to introduce a new technique for the assessment of overt social behavior. This technique, the Riverside Behavioral Q-sort (RBQ), provides ratings of a wide range of behaviors that are part of interpersonal interaction, and is focused at a mid-level of

analysis. Although the RBQ is demanding of time and resources, the technique is not as burdensome as some other behavioral coding strategies. When the technique is carefully applied, the ratings derived can be both reliable and valid. Perhaps most importantly, research emanating from our lab over the past several years has demonstrated that behaviors assessed through the RBQ are correlated with a wide range of other important psychological variables. Analyses of behavior as measured by the RBQ have been informative about cross-situational consistency (Funder & Colvin, 1991) and the consequences of social behaviors associated with inflated self-esteem (Colvin, Block & Funder, 1995), anxiety (Creed & Funder, 1997), unhappiness (Furr & Funder, 1998) and extraversion (Eaton & Funder, 1999). Basic attributes of the RBQ have been described in each these articles. The purpose of the present article is to provide a more detailed description of the technique and its philosophical basis, specific information about its reliability, illustrations of its capacity to reflect both situational and personality effects, and the complete set of items for possible use by other researchers.

Criteria for Development of the Riverside Behavioral Q-Sort

The Riverside Behavioral Q-Sort was originally designed as a scheme for coding videotaped behavioral interactions. The goal was to capture behavior at a level that would not only be psychologically meaningful and relevant to individuals in a behavioral interaction, but that would also require a minimum of subjective interpretation on the part of coders and thereby achieve a sufficient degree of reliability. A balance was sought between molecular, objective units that might be more reliable but less clearly related to psychological phenomena, and more generalized approaches that might be more clearly psychologically meaningful but would also be more subjective and potentially less reliable.

To accomplish these goals, a list of 62 behavioral items was created, the form of a "Q-set" (Funder & Colvin, 1991). A Q-set is a set of descriptive items, typically printed on cards, that raters evaluate by sorting them into a categorical distribution according to how well they characterize whoever or whatever they are being used to describe. Based on several early studies, this Q-set was evaluated and revised. Several items were replaced and the total set expanded to a total of 64 items, yielding the instrument that is employed in our current research

(e.g., Furr & Funder, 1998). The items of the current version, which we now call the Riverside Behavioral Q-sort, are presented in Table 1. Several issues have been carefully considered throughout the development of the RBQ.

Mid-level of analysis. Behavioral analysis can be conducted at several different levels of specificity, ranging from the specific and molecular to the more impressionistic and molar (Mischel, 1973). Each level provides different and potentially important information about what individuals do, so the particular level most appropriate to a given study depends on the kind of information that is needed (Bakeman & Gottman, 1980; Cairns & Green, 1979).

Table 1
RBQ Item Interjudge Agreement and Composite Reliability

RBQ Item		Unstructured Situation		Competitive Situation	
		Avg *r*	Rel	Avg *r*	Rel
1	Expresses awareness of being on camera or in an experiment (regardless of whether reaction is positive or negative).	.46	.77	.29	.62
2	Interviews his or her partner(s) (e.g., asks a series of questions).	.38	.71	.04	.14
3	Volunteers a large amount of information about self.	.29	.62	.20	.50
4	Seems interested in what partner(s) has to say.	.36	.69	.24	.56
5	Tries to control the interaction (disregard whether attempts at control succeed or not).	.17	.46	.25	.58
6	Dominates the interaction (disregard intention, e.g., If subject dominates interaction "by default" because the partner(s) does very little, this item should receive high placement).	.34	.68	.35	.68
7	Appears to be relaxed and comfortable.	.31	.64	.17	.45

Table 1
Continued

RBQ Item		Unstructured Situation		Competitive Situation	
		Avg *r*	Rel	Avg *r*	Rel
8	Exhibits social skills (e.g., does things to make partner(s) comfortable, keeps conversation moving, entertains or charms the partner(s)).	.36	.69	.34	.67
9	Is reserved and unexpressive (e.g., expresses little affect; acts in a stiff, formal manner).	.50	.80	.58	.85
10	Laughs frequently (disregard whether or not laughter appears to be "nervous" or genuine).	.40	.73	.47	.78
11	Smiles frequently.	.29	.62	.36	.69
12	Is physically animated; moves around a great deal.	.23	.55	.25	.57
13	Seems to like partner(s) (e.g., would probably like to be friends with partner(s)).	.31	.64	.19	.49
14	Exhibits an awkward interpersonal style (e.g., seems to have difficulty knowing what to say; mumbles; fails to respond to partner's conversational advances).	.36	.69	.44	.76
15	Compares self to other(s) (whether others are present or not).	.11	.33	.05	.17
16	Shows high enthusiasm and a high energy level.	.40	.73	.49	.79
17	Shows a wide range of interests. (e.g., talks about many topics).	.15	.41	.00	.00
18	Talks at rather than with partner(s) (e.g., conducts a monologue, ignores what partner(s) says).	.23	.54	.21	.52

Table 1
Continued

RBQ Item		Unstructured Situation		Competitive Situation	
		Avg *r*	Rel	Avg *r*	Rel
19	Expresses agreement frequently (high placement implies agreement is expressed unusually often, e.g., in response to each and every statement partner(s) makes. Low placement implies unusual lack of expression of agreement).	.31	.64	.14	.40
20	Expresses criticism (of anybody or anything; low placement implies expresses praise).	.26	.58	.18	.48
21	Is talkative (as observed in this situation).	.38	.71	.45	.77
22	Expresses insecurity (e.g., seems touchy or overly sensitive).	.18	.46	.20	.51
23	Shows physical signs of tension or anxiety (e.g., fidgets nervously, voice wavers). (Lack of signs of anxiety = middle placement; low placement = lack of signs under circumstances where you would expect to see them.)	.21	.51	.13	.38
24	Exhibits a high degree of intelligence (NB: At issue is what is displayed in the interaction, not what may or may not be latent. Thus give this item high placement only if subject actually says or does something of high intelligence. Low placement implies exhibition of low intelligence; medium placement = no information one way or another).	.18	.48	.20	.50
25	Expresses sympathy toward partner(s) (low placement implies unusual lack of sympathy).	.29	.62	.34	.67
26	Initiates humor.	.24	.55	.33	.66
27	Seeks reassurance from partner(s) (e.g., asks for agreement, fishes for praise).	.14	.39	.16	.43

Table 1
Continued

RBQ Item		Unstructured Situation		Competitive Situation	
		Avg *r*	Rel	Avg *r*	Rel
28	Exhibits condescending behavior (acts as if self is superior to partner(s) in one or more ways. Low placement implies acting inferior).	.13	.38	.24	.55
29	Seems likable (to other(s) present).	.27	.59	.23	.54
30	Seeks advice from partner(s).	.22	.52	.33	.66
31	Appears to regard self as physically attractive (nonverbal cues probably will be used to judged this item; examples might include preening, posing, etc.).	.17	.45	.22	.53
32	Acts irritated.	.27	.59	.41	.73
33	Expresses warmth (to anyone, e.g., include any references to "my close friend," etc.).	.21	.52	.19	.48
34	Tries to undermine, sabotage, or obstruct (either the experiment or partner(s)).	.05	.17	.07	.24
35	Expresses hostility (no matter toward whom or what).	.16	.43	.16	.43
36	Is unusual or unconventional in appearance.	.14	.40	.14	.40
37	Behaves in a fearful or timid manner.	.40	.73	.39	.72
38	Is expressive in face, voice, or gestures.	.20	.49	.36	.69
39	Expresses interest in fantasy or daydreams (low placement only if such interest is explicitly disavowed).	.02	.08	.02	.07
40	Expresses guilt (about anything).	.04	.15	.07	.24
41	Keeps partner(s) at a distance, avoids development of any sort of interpersonal relationship (low placement implies behavior to get close to the partner(s)).	.24	.56	.37	.70

Table 1
Continued

RBQ Item		Unstructured Situation		Competitive Situation	
		Avg *r*	Rel	Avg *r*	Rel
42	Shows interest in intellectual or cognitive matters (e.g., by discussing an intellectual idea in detail or with enthusiasm).	.26	.59	.08	.25
43	Seems to enjoy the interaction.	.37	.60	.42	.74
44	Says or does interesting things in this interaction.	.10	.31	.13	.37
45	Says negative things about self (e.g., is self-critical; expresses feelings of inadequacy).	.20	.50	.37	.70
46	Displays ambition (e.g., passionate discussion of career plans, course grades, opportunities to make money).	.36	.69	.01	.04
47	Blames others (for anything).	.08	.27	.04	.14
48	Expresses self-pity or feelings of victimization.	.09	.28	.07	.23
49	Expresses sexual interest (e.g., acts attracted to partner; expresses interest in dating or sexual matters).	.09	.29	.05	.19
50	Behaves in a cheerful manner.	.35	.68	.47	.78
51	Gives up when faced with obstacles (low placement implies unusual persistence).	.17	.46	.31	.64
52	Behaves in a stereotypical masculine/feminine style or manner (apply the usual stereotypes appropriate to the subject's sex. Low placement implies behavior stereotypical of the opposite sex).	.12	.35	.19	.49
53	Offers advice.	.29	.62	.39	.72
54	Speaks fluently and expresses ideas well.	.23	.54	.10	.32
55	Emphasizes accomplishments of self, family, or housemates (low placement = emphasizes failures of these individuals).	.18	.46	.06	.19

Table 1
Continued

RBQ Item		Unstructured Situation		Competitive Situation	
		Avg *r*	Rel	Avg *r*	Rel
56	Competes with partner(s) (low placement implies cooperation).	.06	.20	.29	.62
57	Speaks in a loud voice.	.30	.63	.28	.61
58	Speaks sarcastically (e.g., says things (s)he does not mean; makes facetious comments that are not necessarily funny).	.16	.43	.19	.49
59	Makes or approaches physical contact with partner(s) (of any sort, including sitting unusually close without touching). (Low placement implies unusual avoidance of physical contact, such as large interpersonal distance.)	.18	.46	.21	.51
60	Engages in constant eye contact with partner(s). (Low placement implies unusual lack of eye contact.)	.35	.68	.04	.14
61	Seems detached from the interaction.	.43	.75	.54	.83
62	Speaks quickly (low placement = speaks slowly).	.22	.52	.04	.15
63	Acts playful.	.16	.43	.34	.67
64	Partner(s) seeks advice from subject.	.26	.58	.28	.61

Note. Avg *r* = Average pairwise correlation among four coders. Rel = reliability estimate for the four-coder composite.

At the molecular end of the continuum, some investigators have focused on concrete facial, bodily, and gestural behaviors, and vocal characteristics. Included are variables such as head nods, eyebrow flashes, body orientation, backward lean, sighing, and voice volume (Ekman & Friesen, 1978; Ellgring, 1989; Kalbaugh & Haviland, 1994). At the molar end of the continuum, behavior has been characterized more impressionistically in terms of its broad pattern, style or consequences. For example, the Interpersonal Check List (ICL; La Forge & Suzek, 1955) and the Inventory of Interpersonal Problems (IIP; Horowitz, 1979) have been used to describe the overall behavior of research participants

along broad scales including "managerial-autocratic," "blunt-aggressive" (ICL), and "cold and socially avoidant" (IIP) (e.g., Alden & Phillips, 1990; Hokanson, Lowenstein, Hedeen, & Howes, 1986).

Useful coding schemes can be developed at any point along the molecular versus molar continuum. Our aim was to capture social behavior somewhere near the midpoint. The mid-level of analysis is less concrete and specific than investigations of nonverbal behavior, for example, but is more concrete and specific than ratings of a person's overall social style. Mid-level descriptions might characterize the degree to which an individual in a given situation is talkative, is humorous, or expresses interest in what his or her partner is saying. Such behavioral descriptions could subsequently be extended in either direction along the analytic continuum. They could be broken down into more minute pieces near the molecular end, or combined with other related behaviors to form broader, more molar indicators of behavioral style.

Following the principle that the level of generality at which an investigator aims his or her coding scheme should align closely with the ultimate phenomenon of interest (Bakeman & Gottman, 1980), two considerations motivated our choice of the level at which the RBQ was designed. We hoped ultimately to contribute to an increase in knowledge about the twin issues of (a) how personality is manifested in behavior and (b) how people make inferences about others' personalities from the behaviors they observe (Funder, 1991, 1995, 1999).

In regard to the first consideration, we endeavored to measure behaviors that might be relevant to the aspects of personality described by Block's California Q-set (Block, 1978; the adult version is known as the California Adult Q-set or CAQ). The CAQ is a widely used and comprehensive instrument (McCrae, Costa & Busch, 1986) that assesses 100 mid-level attributes of personality such as "is critical and skeptical," "is genuinely dependable and responsible," and "has a wide range of interests." Notice how these personality descriptors are much broader than specific habits but more specific than broad traits of personality such as the "Big Five" (Goldberg, 1993). Investigations aimed at this mid-level have often painted very rich and detailed portraits of a variety of important psychological phenomena (e.g., Bem & Funder, 1978; Funder, Block, & Block, 1983; Gjerde, Block, & Block, 1988). In a parallel fashion, we attempted to capture behaviors broader (and perhaps more meaningful) than specific responses to individual stimuli, but more specific than general aspects of style. By matching the RBQ level of

analysis to that which has proved so useful with the CAQ, we hoped to create a behavioral measure well aimed at important personality phenomena.

A focus on a mid-level of generality is also compatible with modern answers to the question, "what is a behavior?" While some psychologists seem still to carry a view of "behavior" that dates back to the classical behaviorists, in which each behavior (such as a bar press) was operationally defined at a low and concrete level, modern behaviorism has moved beyond that limited approach. This change in focus has been well described by Walter Mischel (1973, p. 268):

> . . . recent versions of behavioral theory, moving from cat, rat, and pigeon confined in the experimenter's apparatus to people in exceedingly complex social situations, have extended the domain of studied behavior much beyond motor acts and muscle twitches; they seek to encompass what people do cognitively emotionally, and interpersonally, not merely their arm, leg, and mouth movements. Now the term "behavior" has been expanded to include virtually anything that an organism does, overtly or covertly, in relation to extremely complex social and interpersonal events [which] . . . involves inferences about the subject's intentions and abstractions about behavior, rather than mere physical description of actions and utterances.

In regard to the second consideration, that of behavior's role in person perception, we aimed to develop a measure of behavior that matched the phenomenology of lay judges of personality. Although specific evidence on this point is lacking (and should be developed in future research), we believe that the most salient aspect of the phenomenology of behavior—those aspects that ordinary observers notice and utilize in their inferences about personality—are at the mid-level of generality. For example, although "amount of arm swing" while walking has been found to be significantly associated with judgments of walkers' pretended level of sad emotion (Montepare, Goldstein, & Clausen, 1987), individuals in actual social interactions may not ordinarily be consciously aware of such molecular-level behavioral expressions. At the opposite end of the continuum, individuals involved in an interaction may not be consciously aware of the extent to which their partners are being "managerial-autocratic." While we are not denying that a partner's degree of "arm swing" or "autocraticity" may have strong influences on an individual in a social interaction, we feel that it is unlikely that such behaviors are an explicit part of the

individual's conscious awareness. If this assumption is correct, then it may be the mid-level of analysis—such as "tries to control the interaction"—that most closely corresponds to the ordinary phenomenology of an individual engaged in social interaction.

Situational generality. A researcher may be interested in a specific set of behaviors or perhaps only those behaviors that are relevant to specific situations. The coding scheme developed by this researcher is likely to be focused on providing information about only those behaviors. Alternatively, a researcher may be interested in investigating a range of behaviors across several different situations and thus opt for a more general coding system that reflects a wider variety of relevant behaviors.

As is the case when choosing a level of analysis, the most appropriate choice for the level of situational specificity for a coding system is the one that most effectively addresses the issues of interest to the researcher. The goal of the RBQ was to describe behaviors that are generally relevant to many kinds of social interactions. The RBQ has been used to describe individuals' behavior in unstructured, cooperative, and competitive situations, and, although primarily designed for two-person interactions, a modified form for use with group discussions is under development. So far, all of these settings have been videotaped encounters held within our lab, but many of the items are general enough to be relevant to other kinds of social interaction. For example, items such as "Is talkative" and "Expresses criticism" are potentially applicable to almost any social situation. This flexibility affords the potential for tracking the consistency and distinctiveness of behavior across many different situations and interactions, an ability that may be useful for investigations on a wide range of topics.

Our interest in the relationships between personality and behavior demanded an instrument that would have clear relevance for a wide range of important personality characteristics. Once again, the CAQ proved to be a valuable guide. Thirty-nine RBQ items were deliberately written to correspond closely with CAQ items. The CAQ items chosen for RBQ development were personality characteristics with clear relevance for specifiable behaviors in dyadic interactions. For example, CAQ item 92, "Has social poise and presence; appears to be socially at ease" is directly transferable to an observable behavior (RBQ item 8, "Exhibits social skills"), whereas CAQ item 45, "Has a brittle ego-defense system; has a small reserve of integration; would be disorganized and maladaptive

when under stress or trauma," has less direct behavioral relevance for normal dyadic situations. The correspondences between RBQ and CAQ items are purposely not subtle. The intention was to write behavioral items that directly characterized surface manifestations of important personality traits; more complex or deep relationships—to the extent any exist—must be discovered through other means.

Observability. When writing the items for the RBQ, we deliberately restricted ourselves to descriptions of *overt* behavior. That is, the descriptions purposely focus on what behaviors *are* (or superficially appear to be), and to avoid inferences about what the behaviors *mean*. For example, when coding the item "laughs frequently," coders are explicitly told not to attempt to distinguish between "nervous" and "genuine" laughter. The occurrence of laughter is the phenomenon the technique is intended to detect; its meaning must be assessed through other methods. Similarly, items referring to emotional expression describe the emotion that appears to be expressed, while avoiding any inferences about what the underlying emotion might really be. The item "Seems interested in what partner has to say" refers to the subject's apparent interest, as expressed overtly. It does not have to be—indeed it should not be—rated with respect to whether the coder believes the expression of interest to be genuine or feigned. Items such as "Appears to be relaxed and comfortable" are rated in a similar fashion. It is the overt behavioral expression, not the underlying affect, that is rated. These aspects of the phrasing of the items, along with instructions to coders to rate what they see, rather than what they infer, are meant to maximize the ratings' objectivity, interjudge reliability, and interpretability.

This emphasis on the observability and surface characteristics of the behaviors that are rated does not mean that we are uninterested in deeper aspects of personality and underpinnings of behavior. It just means we are skeptical about the utility of asking behavioral coders to divine what these are. The true meaning of these behaviors must emerge in other ways, such as an examination of the ways in which they are correlated with personality, phenomenology, or physiology.

Procedural feasibility. The RBQ was designed to be relatively simple for coders to learn and use. Some other techniques require the coding of intervals that last for a few seconds or less (e.g., Youngren & Lewinsohn, 1980) or the identification of complex and subtle movements (Ekman & Friesen,

1978). Techniques like these require extensive—and sometimes expensive—training of coders and may be open to misuse or confusion on the part of coders who are less than expert. The complexity involved in these kinds of systems may be required in order to address fully the questions posed by the researchers who developed them, but the RBQ was specifically designed to avoid such complexity. As anyone who has attempted to do any systematic behavioral coding can attest, the process always requires a significant investment of time and other resources. Our intention was to make the process as simple as possible, while ensuring that it remained a useful and valid system of description.

To accomplish these goals, the RBQ was designed to rely heavily on the observational ability and common sense of the coders. Any socially competent individual has experience in many different contexts and is able to make reasonably accurate judgments about the extent to which an individual appears to be irritated, expressive, or relaxed. Indeed, despite evidence that human judges are susceptible to certain inferential errors (e.g., Kahneman & Tversky, 1973; Nisbett & Ross, 1980), further research has demonstrated the considerable ability of human beings to make accurate and effective inferences about a variety of phenomena (Funder, 1995, 1999; Hammond, 1996). Thus, rather than requiring coders to mechanically tally the occurrences of some set of behaviors or record time spent in specified activities, the RBQ process asks coders to watch an interaction and then estimate the extent to which a variety of behaviors were relatively characteristic of the focal participant. This procedure assumes that human coders can describe with reasonable accuracy the extent to which an individual initiates humor, dominates an interaction, and seems to like his or her partner. The test of this assumption lies in the quality—the reliability and validity—of the data that emerge.

The relative simplicity of the coding procedure—observing an interaction and rating each of the RBQ items—is intended to make the process manageable. Furthermore, the simplicity also corresponds well with the first goal considered for the RBQ, which is to focus on a middle level of analysis. All items of the RBQ describe behaviors that are familiar to almost any socially experienced adult. Therefore, the coding scheme does not require a great amount of explanation or training for coders to understand the units of analysis. Typically, two 2-hour training sessions are sufficient. No particular apparatus is required beyond the Q-sort deck

itself, a sheet on which RBQ descriptions can be recorded, and the video equipment with which coders typically view the behaviors.

Format. The 64 RBQ items are deployed in the form of a Q-sort deck (Stephenson, 1953). The typical Q-sort consists of a set of cards, each with a different descriptor printed on it. Q-sort items can reflect any kind of construct including personality characteristics, behaviors, attitudes, and social attributes. In using a Q-sort, an individual is asked to rate a particular object with respect to the descriptors printed on the cards. Usually, the cards are placed in a predetermined or "forced" distribution that ranks the extent to which the descriptors are relatively characteristic of the object.

Several excellent sources extensively discuss the properties and advantages of the Q-sort methodology (Block, 1978; Caspi et al., 1992; Ozer, 1993). One significant advantage of the "ipsative" Q-sorting procedure is that the use of a forced distribution (see below) ensures that the ratings of all judges have the same mean and standard deviation, computed across items. This property of the Q-sort reduces the possible effect of various rating response sets such as acquiescence or extremity bias. It also might mitigate, if not eliminate, social desirability biases, because the forced-choice procedure ensures that not all desirable item may be rated high, nor all undesirable items be rated low. Finally, because the procedure requires that each item be placed in relation to every other item, it forces the judge to make more, finer, and perhaps more carefully considered distinctions than typical rating methods.

In our procedures, judges focus on a single participant during a behavioral interaction, then place the cards into a quasi-normal distribution of nine categories that range from "not at all or negatively characteristic of the behavior of the person (1)" to "highly characteristic of the person's behavior (9)." Behaviors evaluated as neither characteristic or uncharacteristic are placed in the middle category (5).

To ensure the form of the distribution, judges are asked to place a predetermined number of cards into each category, specifically for categories 1 through 9: 3, 5, 7, 10, 14, 10, 7, 5, and 3. This quasi-normal distribution is similar to that usually used for the CAQ, personality Q-sort (both sorts place approximately 5% of their items in each extreme category, causing a placement of an item into either category to be a strongly implicative and difficult choice). While it would be possible to have coders rate each RBQ item on an unforced, Likert scale rather than

sort them into a forced distribution, this would change the instrument's ipsative properties described above. The extent to which this change, or changes in the preassigned distribution frequencies, would alter the measure's ability to reflect the content of interpersonal interactions is an empirical question that deserves further research.

METHOD

We have used the RBQ to code the behavior of individuals in social situations that were set up and videotaped within our laboratory. The specific procedures we followed when assembling our current major data set (the Riverside Accuracy Project), and findings concerning reliability and validity are summarized below.

Participants

A total of 184 undergraduate participants (92 female, 92 male) were recruited by the Riverside Accuracy Project (Funder, 1995). Participants engaged in an extensive variety of tasks, including providing personality descriptions of themselves and engaging in three behavioral interactions. Due to occasional technical difficulties and subject attrition across sessions, the actual sample size for any given analysis is smaller than the total and varies slightly from one analysis to the next. All participants were paid for their time.

Procedures

Social interactions. Participants engaged in a series of three dyadic interactions with an opposite-sex stranger (one of the other participants). Each interaction was approximately 5 min long and was recorded by a video camcorder set up in plain sight of both participants. The situations in which these interactions took place were designed to provide realistic interpersonal encounters that varied across some of the dimensions that differentiate situations in real life. Specifically, one situation was designed to provide opportunities for cooperation, one to provide opportunities for competition, and another was designed to provide as little structure and as much behavioral latitude as possible. Each situation was conducted with a randomly assigned, previously unacquainted partner of the opposite sex, to capture a bit of how participants behave in cross-gender interactions.

In the first, unstructured interaction, participants were simply seated on a couch and encouraged to talk about whatever they would like. In the second, cooperative interaction, they were seated together at a table and given 5 min to

build a tinker toy that matched a model; each received $1 if they succeeded. In the third, competitive interaction, they played the popular sound-repetition game, Simon®. The winner of three games out of five was paid an extra $1. In the interest of brevity, we will focus mainly on these unstructured and competitive interactions with opposite-sex strangers in our exploration of the important qualities of the RBQ. Parallel data, however, concerning the remaining, cooperative situation are posted on the World Wide Web.[1]

Sorting and quality control. To obtain reliable descriptions of the behaviors manifest on these tapes, we acquired four codings of *each participant in each situation*. Undergraduate research assistants were trained in use of the RBQ,[2] independently watched assigned videotaped interactions, and provided RBQ descriptions of participants. Each research assistant coded many different participants, but viewed only one interaction for any given participant. In addition, coders were instructed to disregard a coding assignment and notify the research supervisor if they had any acquaintance with an assigned participant. These procedures were designed to ensure that the description of a participant was based only on the behavior displayed by the participant in the interaction being coded and was not influenced any previous observation of that individual.

Coders were carefully instructed in the use of the 9-point categorical rating system. For example, as was mentioned earlier, they were instructed to concentrate on the observable aspects of the behavior they rated, and to avoid inferences about what behaviors "really" were and meant. In our experience, as coders begin to formulate elaborate explanations for their ratings, instead of just rating what they see, they become less reliable (agreeing less with other coders) and very probably less valid as well. Coders received no explicit instruction regarding the *content* of the items. Rather than impose our own rigid definitions, it seemed better to allow coders to use their own common sense in identifying dominance behaviors, relaxation behaviors, or social skill behaviors (to name just three items).

1. The address for additional data and analyses from this study is http://www.psych.ucr.edu/faculty/funder/rap/Supplemental/RBQ.pdf.

2. A copy of the instruction booklet associated with our training procedures is presented in Appendix A. Training consisted of several tasks, including full descriptions of the Q-sorting procedure, instructions regarding observation of interactions and judgments regarding RBQ items (e.g., no discussion with other coders, limiting the degree of inferences about possible underlying traits, motivations, and feelings), discussions of the importance of confidentiality in research, and instructions about practical issues involving access to videotapes, monitors, and so forth. In addition, potential coders performed a trial RBQ description of a specific target participant. Each potential coder's description was compared to criterion RBQ descriptions of the specified target provided by several graduate students and faculty in psychology. Coders whose description correlated well with the criterion descriptions were recruited for extensive coding work.

Ongoing checks ensured that only codings that met certain criteria were retained. Agreement among coders was continually assessed by computing profile-level correlation coefficients between the 64 ratings provided by each coder and the 64 ratings provided by other coders that became available as they completed their work. When four codings were obtained, each coding had to agree with two other codings at least $r = .30$ and one other at least $r = .20$. If it did not, the coding was deleted and assigned to another coder to complete.

We used this ongoing and preliminary method of assessing interjudge agreement primarily as a means of quality control. We have found the method useful for detecting particular instances where coders were tired, inattentive, or misunderstanding of the procedure. In such cases, a very low or even negative profile correlation between one coder and the others for that session allows us to detect and correct a mistaken coding. These profile correlations, however, are poor indicators of interjudge reliability for any other purpose, being strongly affected by "stereotype accuracy" and other potential confounds identified by Cronbach (1955).

For this reason, our formal assessment of reliability concentrates on single items. It is also appropriate to concentrate on the reliabilities of single items because most of our research has been based on collections of correlations between single behavioral items and predictor variables such as personality trait scores. Single-item reliabilities can be computed only after coding is completed for multiple subjects in a sample. For each participant in each situation, the average was computed for each of the 64 RBQ items across the four coders. These 64 averaged codings were then employed as the description of the individual's behavior in that particular situation. The outcomes of the reliability analyses are summarized in the Results section.

Personality ratings. Participants completed personality descriptions of themselves using, among other instruments, the NEO-Personality Inventory (NEO-PI; Costa & McCrae, 1985), a well-validated measure of the "Big Five" personality traits, and the Beck Depression Inventory (BDI; Beck, Ward, Mendelsohn, Mock, & Erbaugh, 1961), the most widely used inventory of depressive affect.

RESULTS

Reliability Analyses

The first step in our analysis of the RBQ was to assess the reliability of the four-coder composite score for each item. As mentioned in the previous section, a different panel of judges rated each of the three situations for each target, in order to keep estimates of cross-situational

consistency uncontaminated by raters' memory for how targets acted in other situations.

Calculating reliabilities. The full content of the RBQ items, and their reliabilities in two sessions (the unstructured and competitive) are presented in Table 1.[3] This table includes the average pairwise correlation for each item and the reliability of the four-coder composite (see Shrout & Fleiss, 1979, equations ICC (1,1) and ICC (1,k)). The predicted reliability for panels of sizes other than four can be readily computed using the average pairwise correlation and the Spearman-Brown formula.

The reliability estimates in the unstructured situation ranged from .08 to .80, with a mean of .53. Thirty-nine of the RBQ items had estimated reliabilities above .50 and only one had an estimated reliability below .10. Reliabilities in the competitive situation were similar, with a range from .00 to .85 and a mean of .50. Thirty-five of the items had estimated reliabilities higher than .50 and three were below .10. The items with very low reliabilities typically referred to behaviors rarely exhibited in these situations (e.g., item 39, "expresses interest in fantasy and daydreams"). Such items are retained in the RBQ to retain a flexibility of application across other situations in which different behaviors may be salient (Furr & Funder, 1998).

Evaluating reliabilities. Are these reliabilities acceptable? When evaluating the figures in Table 1, three points should be kept in mind.

First, the rigorous procedure by which we assigned coders to targets required that no coder rate or see the same target in more than one situation. Because each interaction contained two participants, no coder could rate more than half the targets in a given interaction. This constraint, along with other practicalities of coder assignment (e.g., the maximum number of assignments that could be given to each coder), while crucial for the validity of our investigations of cross-situational consistency, also had the effect of making reliable differences among coders inseparable from reliable differences among participants. The presently obtained reliabilities may therefore be lower than those that would emerge from a

3. These two sessions are the ones across which, in validity analyses, we will present data concerning behavioral consistency and change. The reliabilities in the remaining, cooperative situation are quite similar and can be viewed at our Web site.

design, perhaps one unconcerned with estimates of cross-situational consistency, that imposed fewer constraints.

Second, the RBQ is related to the personality-descriptive CAQ. Not only were many items on the RBQ inspired by the content of the CAQ, but both instruments approach psychological constructs at a middle level of abstraction, and both are designed to be used with judges or observers. The CAQ has a long and distinguished history of successful application, so it is natural to ask this question: How does the reliability of the RBQ compare to that of the CAQ?

This question can be answered within our data. Two close acquaintances of each participant, and two strangers who viewed the participant for only 5 minutes, described him or her with the CAQ. To put the reliabilities of the RBQ, presented in Table 1, on the same scale as those of the CAQ, one can use the Spearman-Brown prophecy formula to estimate the reliability of each CAQ item if each had been rated by a panel of four acquaintances, or four strangers. Recall that the average RBQ item reliabilities in the two situations were .53 and .50; the comparable average four-judge composite reliabilities for the CAQ were .32 when the judges were strangers, and .45 when the judges were acquaintances. Thus, the RBQ appears to have the ability to generate reliabilities comparable to those of the widely used CAQ.

A third point is the most important of all. Statistical textbooks typically present benchmarks for acceptable reliabilities that are quite high—the figure of .80 is not uncommon—and in some contexts rather unrealistic. The origin of these benchmarks is seldom made explicit. They are typically presented with great confidence but without explicit justification, and a reader might be forgiven for suspecting they are arbitrary. So it is worth asking, what is the ultimate purpose of reliable measurement?

The answer is quite obvious: If the reliability of a measurement is high, the chances of it generating meaningful relations with other measures of psychological phenomena, and therefore of it being psychologically informative and predictively useful, is also high. In other words, reliability is sought to improve the chances of attaining validity and in the end it is validity that matters. For example, the widespread use of the CAQ stems from the validity it has demonstrated in many contexts, not demonstrations of its unfailing high reliability. In the same way, we suggest that the best way to evaluate the meaningfulness of the RBQ is to consider its validity, the meaningful relations it can generate with other

measurements and the light it can shed on psychological phenomena. Such relations are the topic of the next section.

Validity Analyses

The validity of the RBQ for capturing important aspects of social behavior can be demonstrated in two ways. First, it can be used to detect differences between situations in their effects on the average behavior of research participants. Second, it can be used to detect relationships between personality-relevant individual difference and behavior.

Situational effects. The experiment is the traditional method of social psychology for demonstrating the effect of situations on behavior. Participants are exposed to situations that differ along one or more dimensions of interest, and a behavioral dependent variable is measured. The RBQ offers a technique for measuring 64 behavioral dependent variables at once. By showing how these behaviors change across situations, we can demonstrate the sensitivity of RBQ-measured behaviors to situational variables, and also reveal behaviorally consequential differences between situations.

Tables 2 and 3 apply this technique to a comparison of two conditions in our study, the unstructured and competitive interaction, reported separately for each sex. It can immediately be seen that, considered as a repeated measures experiment, the results are strong and informative. Fully 46 out of 64 behavioral dependent variables were significantly different (at $p < .05$) between the two situations among female participants; the proportion was 44 out of 64 among male participants.

In the unstructured interaction, participants were relatively talkative on a wide range of topics, warm, sociable, agreeable, and fluent, compared to the competitive situation. In the competitive interaction, they were relatively competitive (not surprisingly), playful, enthusiastic, loud, animated, and tending to smile and laugh more often, compared to the unstructured situation.

These results demonstrate two things. First, they show that the RBQ is finely sensitive to the wide-ranging effects of situational differences on behavior. Second, they are informative about the *nature* of these situational differences. The unstructured interaction can be characterized as a situation that elicits a polite, friendly, and rather low-key conversation. The competitive interaction can be characterized as a situation that

Table 2
Mean Differences in RBQ Between Unstructured and Competitive Situations—Females

	RBQ Factor/Item	Unstructured M	Competitive M	t
Items with higher means in the unstructured situation				
2	**Interviews his or her partner(s)**	**7.31**	**3.81**	**22.24*** **
60	**Engages in constant eye contact**	**7.19**	**4.31**	**18.43*** **
3	**Volunteers a large amount of information**	**6.62**	**3.60**	**18.09*** **
17	**Shows a wide range of interests**	**5.93**	**4.17**	**13.72*** **
21	**Is talkative**	**7.05**	**5.10**	**9.66*** **
46	**Displays ambition**	**6.05**	**4.82**	**8.20*** **
42	**Shows interest in intellectual matters**	**5.66**	**4.84**	**8.01*** **
8	**Exhibits social skills**	**7.44**	**6.23**	**7.78*** **
4	**Seems interested in what partner(s) says**	**7.00**	**5.88**	**7.70*** **
	Confidence (factor)	**5.52**	**5.07**	**6.94*** **
54	**Speaks fluently and expresses ideas well**	**6.83**	**6.09**	**5.51*** **
33	**Expresses warmth**	**5.62**	**5.02**	**5.26*** **
19	**Expresses agreement frequently**	**5.99**	**5.43**	**4.51*** **
55	**Emphasizes accomplishments**	**5.45**	**4.89**	**4.47*** **
	Positive Affect with Partner (factor)	**6.20**	**5.41**	**4.44*** **
51	**Gives up when faced with obstacles**	**4.60**	**4.02**	**3.91*** **
24	**Exhibits a high degree of intelligence**	**5.27**	**4.95**	**3.44*** **
20	**Expresses criticism**	**5.36**	**4.88**	**3.37** **
1	**Aware of being on camera or in experiment**	**5.85**	**5.31**	**2.47***
39	**Expresses interest in fantasy or daydreams**	**5.02**	**4.87**	**2.33***
47	Blames others (For anything)	4.24	4.01	2.05*

Table 2
Continued

RBQ	Factor/Item	Unstructured M	Competitive M	t
	Items with higher means in the competitive situation			
56	**Competes with partner(s)**	**3.60**	**6.14**	**–13.25*** **
63	Acts playful	4.09	5.70	–10.40***
16	**Shows high enthusiasm and energy level**	**4.89**	**6.27**	**–8.50*** **
10	**Laughs frequently**	**6.28**	**7.59**	**–7.82*** **
11	**Smiles frequently**	**6.98**	**7.87**	**–7.31*** **
57	**Speaks in a loud voice**	**4.41**	**5.33**	**–6.54*** **
30	**Seeks advice from partner(s)**	**4.48**	**5.45**	**–6.51*** **
12	**Is physically animated**	**4.19**	**5.20**	**–6.13*** **
18	**Talks at rather than with partner(s)**	**2.88**	**3.63**	**–5.77*** **
38	**Is expressive in face, voice or gestures**	**5.98**	**6.97**	**–5.75*** **
27	**Seeks reassurance from partner(s)**	**4.47**	**5.17**	**–5.54*** **
59	**Makes or approaches physical contact**	**3.88**	**4.65**	**–5.42*** **
50	**Behaves in a cheerful manner**	**6.70**	**7.45**	**–5.12*** **
53	**Offers advice**	**4.62**	**5.34**	**–4.36*** **
64	**Partner(s) seeks advice from subject**	**4.46**	**5.11**	**–4.29*** **
62	**Speaks quickly**	**4.96**	**5.46**	**–4.26*** **
43	**Seems to enjoy the interaction**	**6.58**	**7.30**	**–4.11*** **
45	**Says negative things about self**	**3.85**	**4.59**	**–3.90*** **
52	**Behaves stereotypically feminine**	**5.80**	**6.24**	**–3.47*** **
22	Expresses insecurity	3.72	4.22	–3.37**
13	**Seems to like partner(s)**	**5.96**	**6.31**	**–3.32** **
14	Exhibits an awkward interpersonal style	3.18	3.67	–2.59*
29	**Seems likable. (To other(s) present)**	**6.55**	**6.83**	**–2.39***
7	**Appears to be relaxed and comfortable**	**6.74**	**7.20**	**–2.35***
32	Acts irritated	2.99	3.38	–2.27*

Table 2
Continued

RBQ	Factor/Item	Unstructured M	Competitive M	t
5	Tries to control the interaction	3.86	4.24	–2.11*
	Involvement (factor)	**5.85**	**6.02**	**–2.11***
26	**Initiates humor**	**4.95**	**5.27**	**–2.09***

Note. $N = 82$. Item content abbreviated. *** = $p < .001$. ** = $p < .01$. * = $p < .05$. Bold items are also significant in male sample (see Table 3). All tests were two-tailed.s were two-tailed.

elicits playful, energetic, loud, and enthusiastic game playing. In future research, as the RBQ is applied to an ever-increasing variety of situations, it offers the potential to be the basis of a method for comparing the general effects of different situations on behavior, and perhaps eventually become helpful in characterizing the "personality of situations" (Bem & Funder, 1978).

Personality effects. The correlational study is the traditional method of personality psychology for demonstrating the relationships between personality and behavior. Participants with a range of scores on a personality dimension are compared with one another as to the degree they attain a criterion or exhibit a behavior of interest. With the RBQ, 64 different behaviors can be associated with each personality score. The assessment of these correlations can demonstrate the convergent validity of the RBQ (in terms of its convergence with personality assessments) as well as reveal some of the behavioral concomitants of personality.

For present purposes, we investigated correlates between RBQ-coded behaviors (averaged over the unstructured and competitive interactions considered earlier) and two important dimensions of personality, extraversion and depression (or personal negativity; see Furr & Funder, 1998).[4] Extraversion was assessed with the NEO-PI (Costa & McCrae, 1985)

4. These results are presented because they were particularly interesting and serve to demonstrate the validity of the RBQ with respect to these variables. Other analyses using other variables—such as the other four factors assessed by the NEO—can be viewed on our Web site.

Table 3
Mean Differences in RBQ Between Unstructured and Competitive Situations—Males

	RBQ Factor/Item	Unstructured M	Competitive M	t
Items with higher means in the unstructured situation				
2	**Interviews his or her partner(s)**	**6.87**	**3.75**	**18.68*** **
3	**Volunteers a large amount of information**	**6.72**	**3.33**	**18.23*** **
60	**Engages in constant eye contact**	**7.02**	**4.27**	**16.15*** **
17	**Shows a wide range of interests**	**5.86**	**4.16**	**11.50*** **
46	**Displays ambition**	**6.26**	**4.84**	**9.56*** **
21	**Is talkative**	**6.80**	**5.09**	**7.18*** **
42	**Shows interest in intellectual matters**	**5.63**	**4.91**	**6.35*** **
55	**Emphasizes accomplishments**	**5.62**	**4.92**	**6.21*** **
	Confidence (factor)	**5.56**	**5.19**	**5.80*** **
4	**Seems interested in what partner(s) says**	**6.88**	**6.04**	**5.60*** **
33	**Expresses warmth**	**5.40**	**4.89**	**4.09*** **
23	Shows physical signs of tension or anxiety	5.04	4.37	3.92***
8	**Exhibits social skills**	**6.85**	**6.25**	**3.48*** **
54	**Speaks fluently and expresses ideas well**	**6.49**	**6.01**	**3.18***
	Positive Affect with Partner (factor)	**6.08**	**5.85**	**3.18***
20	**Expresses criticism**	**5.54**	**5.04**	**3.18***
51	**Gives up when faced with obstacles**	**4.46**	**3.98**	**2.90***
34	Tries to undermine, sabotage or obstruct	3.38	2.96	2.86**
1	**Aware of being on camera or in experiment**	**6.21**	**5.58**	**2.56***
44	Says or does interesting things	5.75	5.40	2.49*
39	**Expresses interest in fantasy or daydreams**	**4.99**	**4.87**	**2.39***
19	**Expresses agreement frequently**	**5.92**	**5.64**	**2.16***

Table 3
Continued

RBQ	Factor/Item	Unstructured M	Competitive M	t
Items with higher means in the competitive situation				
56	**Competes with partner(s)**	**3.70**	**6.27**	**–12.69*** **
63	**Acts playful**	**3.96**	**5.17**	**–8.02*** **
16	**Shows high enthusiasm and energy level**	**4.22**	**5.55**	**–7.40*** **
64	**Partner(s) seeks advice from subject**	**4.66**	**5.66**	**–6.73*** **
10	**Laughs frequently**	**5.61**	**6.88**	**–6.39*** **
30	**Seeks advice from partner(s)**	**4.33**	**5.25**	**–6.05*** **
38	**Is expressive in face, voice or gestures**	**5.32**	**6.40**	**–5.82*** **
7	**Appears to be relaxed and comfortable**	**6.47**	**7.47**	**–5.74*** **
50	**Behaves in a cheerful manner**	**6.09**	**7.03**	**–5.48*** **
57	**Speaks in a loud voice**	**4.42**	**5.32**	**–5.34*** **
11	**Smiles frequently**	**6.69**	**7.52**	**–5.33*** **
18	**Talks at rather than with partner(s)**	**3.02**	**3.79**	**–5.20*** **
29	**Seems likable. (To other(s) present)**	**6.43**	**6.95**	**–4.75*** **
27	**Seeks reassurance from partner(s)**	**4.40**	**5.02**	**–4.38*** **
62	**Speaks quickly**	**4.81**	**5.40**	**–4.26*** **
53	**Offers advice**	**4.90**	**5.66**	**–4.18*** **
43	**Seems to enjoy the interaction**	**6.48**	**7.24**	**–4.01*** **
12	**Is physically animated**	**4.14**	**4.82**	**–3.74*** **
59	**Makes or approaches physical contact**	**4.00**	**4.47**	**–2.99** **
45	Says negative things about self	3.95	4.38	–2.79**
52	**Behaves stereotypically masculine**	**5.88**	**6.22**	**–2.73** **
28	Exhibits condescending behavior	4.42	4.73	–2.53*
	Involvement (factor)	**5.61**	**5.85**	**–2.50***
13	**Seems to like partner(s)**	**6.20**	**6.52**	**–2.38***
26	**Initiates humor**	**5.13**	**5.50**	**–2.22***

Note. $N = 78$. Item content abbreviated. *** = $p < .001$. ** = $p < .01$. * = $p < .05$. Bold items are also significant in female sample (see Table 2). All tests were two-tailed.

and depression with the BDI (Beck et al., 1961). The results, displayed separately for each sex, are shown in Tables 4–7.[5]

The results for extraversion, seen in Tables 4 and 5, are strong and clear. Thirteen of the 64 RBQ items are significantly associated (at $p < .05$) with this trait among females, and even more, 28 out of 64, are associated among males. In both sexes, extraversion is associated with behavior that is enthusiastic and animated, and with tendencies to be humorous, dominant, socially skilled and forthcoming. Extraversion is *negatively* associated, in both sexes, with behaviors that are insecure, timid, critical, awkward, and anxious, among many more correlates. These results are sensible and help establish the convergent validity of the RBQ with the NEO-PI as well as providing a vivid portrait of the extraverted behavior pattern.

The results for depression, presented in Tables 6 and 7, are equally informative. Whereas extraversion is a broad personality trait, depression is a narrower trait that may be more strongly related to internal mental and emotional experiences than to social activity. Nevertheless, analysis of the RBQ reveals that this trait does indeed have meaningful relationships with social behavior, especially among females, where 12 of the 64 RBQ items are significantly associated with depression. For females, depression is positively associated with behavior that is critical, insecure, self-pitying, and irritated, and negatively associated with verbal fluency, cheerfulness, and comfort. Males, on the other hand, show almost no visible, behavioral expression of depression. These findings correspond well with research demonstrating significant gender differences in the expression of depression (Funabiki, Bologna, Pepping, & Fitzgerald, 1980; Hammen & Padesky, 1977; Nolen-Hoeksema, 1987, 1990; Vredenburg, Krames, & Flett, 1986), and provide more evidence for the convergent validity of the RBQ.

Principal Components of the RBQ

To date, research using the RBQ has focused on the analysis of mean changes in and correlates of individual items, such as we have just seen.

5. The results for NEO-PI Extraversion and the BDI overlap somewhat with results presented elsewhere. Eaton and Funder (1999) have correlated extraversion with RBQ descriptions of unstructured interactions with opposite-sex strangers, and Furr and Funder (1998) correlated BDI scores with RBQ descriptions averaged across all three situations with the opposite-sex stranger.

Table 4
RBQ Scores Averaged Across Two Situations: Correlates of NEO-PI Extraversion—Females

RBQ Factor/Item		r
56	Competes with partner(s)	.32**
26	**Initiates humor**	**.32****
	Confidence (factor)	.29**
	Involvement (factor)	**.28***
57	Speaks in a loud voice	.26*
16	**Shows high enthusiasm and energy level**	**.24***
12	**Is physically animated**	**.23***
44	**Says or does interesting things**	**.22†**
8	**Exhibits social skills**	**.21†**
3	**Volunteers a large amount of information**	**.21†**
54	Speaks fluently and expresses ideas well	.21†
6	**Dominates the interaction**	**.21†**
28	Exhibits condescending behavior	.20†
5	Tries to control the interaction	.19†
22	**Expresses insecurity**	**–.41*****
48	Self pity or feelings of victimization	–.35**
37	**Behaves in a fearful or timid manner**	**–.33****
40	**Expresses guilt**	**–.33****
27	**Seeks reassurance from partner(s)**	**–.26***
20	**Expresses criticism**	**–.25***
14	**Exhibits an awkward interpersonal style**	**–.24***
51	**Gives up when faced with obstacles**	**–.22***
9	**Is reserved and unexpressive**	**–.19†**
30	Seeks advice from partner(s)	–.19†
41	**Keeps partner at a distance**	**–.18†**
	Positive Affect with Partner (factor)	–.07

Note. Item content abbreviated. *** = $p < .001$. ** = $p < .01$. * = $p < .05$. † = $p < .10$. $n = 83$. Bold items are also significant in both genders (see Table 5).

We prefer this use because it seems to provide the most vivid portrayal of what participants are actually doing in various contexts. For some purposes, however, other investigators might wish to rise above this mid-level of analysis and consider composites of RBQ items that have similar content.

Table 5
RBQ Scores Averaged Across Two Situations: Correlates of NEO-PI Extraversion—Males

RBQ	Factor/Item	r
	Involvement (factor)	**.45*** **
63	Acts playful	.42***
43	Seems to enjoy the interaction	.41***
26	**Initiates humor**	**.39*** **
8	**Exhibits social skills**	**.38*** **
21	Is talkative	.38***
13	Seems to like partner(s)	.37***
50	Behaves in a cheerful manner	.35**
16	**Shows high enthusiasm and energy level**	**.32****
38	Is expressive in face, voice or gestures	.31**
44	**Says or does interesting things**	**.30****
	Positive Affect with Partner (factor)	.27*
12	**Is physically animated**	**.26***
29	Seems likable	.23*
2	Interviews his or her partner(s)	.21†
4	Seems interested in what partner has to say	.21†
49	Expresses sexual interest	.20†
7	Appears relaxed an comfortable	.20†
6	**Dominates the interaction**	**.19†**
3	**Volunteers a large amount of information**	**.19†**
	Confidence (factor)	.17
20	**Expresses criticism**	**–.44*** **
61	Seems detached from the interaction	–.42***
51	**Gives up when faced with obstacles**	**–.41*** **
9	**Is reserved and unexpressive**	**–.40*** **
32	Acts irritated	–.39***
41	**Keeps partner(s) at a distance**	**–.37*** **
23	**Shows physical signs of tension or anxiety**	**–.34****
18	Talks at rather than with partner(s)	–.31**
22	**Expresses insecurity**	**–.30****
14	**Exhibits an awkward interpersonal style**	**–.30****
37	**Behaves in a fearful or timid manner**	**–.29****
47	Blames others (for anything)	–.28**
28	Exhibits condescending behavior	–.28*
40	**Expresses guilt**	**–.25***

Table 5
Continued

RBQ	Factor/Item	r
42	Shows interest in intellectual matters	–.22*
35	Expresses hostility	–.20†
27	Seeks reassurance from partner(s)	–.20†

Note. Item content abbreviated. *** = $p < .001$. ** = $p < .01$. * = $p < .05$. † = $p < .10$. $n = 82$; Bold items are also significant in both genders (see Table 4).

Table 6
RBQ Scores Averaged Across Two Situations: Correlates of the BDI — Females

RBQ	Factor/Item	r
48	Self pity or feelings of victimization	.37***
20	Expresses criticism	.36***
32	Acts irritated	.33**
51	Gives up when faced with obstacles	.32**
45	Says negative things about self	.29**
22	Expresses insecurity	.28*
15	Compares self to other(s)	.22*
40	Expresses guilt	.22†
49	Expresses sexual interest	.20†
37	Behaves in a fearful or timid manner	.19†
14	Exhibits awkward interpersonal style	.19†
54	Speaks fluently and expresses ideas well	–.31**
	Confidence (factor)	–.29**
7	Appears to be relaxed and comfortable	–.27*
50	Behaves in a cheerful manner	–.26*
60	Engages in constant eye contact	–.25*
17	Shows a wide range of interests	–.23*
43	Seems to enjoy the interaction	–.20†
	Involvement (factor)	–.20†
8	Exhibits social skills	–.19†
	Positive Affect with Partner (factor)	–.11

Note. Item content abbreviated.* = $p < .05$. ** = $p < .01$. *** = $p < .001$. $n = 81$.

Table 7
RBQ Scores Averaged Across Two Situations: Correlates of the BDI—Males

RBQ	Factor/Item	r
41	Keeps partner(s) at a distance	.25*
28	Exhibits condescending behavior	.21†
	Confidence (factor)	.16
48	Self pity or feelings of victimization	–.27*
44	Says or does interesting thing	–.27*
25	Expresses sympathy toward partner(s)	–.23†
	Positive Affect with Partner (factor)	–.18
	Involvement (factor)	–.10

Note. Item content abbreviated. * = $p < .05$. ** = $p < .01$. *** = $p < .001$.
$n = 65$.

Factor analysis is not technically appropriate for the RBQ or other ipsative instruments because of the dependencies among items produced by the forced-choice rating procedure. A principal components analysis was performed as part of a recent doctoral dissertation (Wiehl, 1997). This analysis focused on the first, unstructured interaction. A three-component solution accounted for 43% of the variance in RBQ items. After a varimax rotation, three components were examined and labeled "involvement," "positive affect," and "confidence." All but eight items had loading above .40 on their main component, and few items loaded on more than one component.

Involvement items included "shows high enthusiasm and a high energy level," "is talkative," and "is reserved and unexpressive" (reversed). *Positive affect* items included "seems interested in what partner says," "seems to like partner," and "talks at rather than with partner" (reversed). *Confidence* items included "exhibits a high degree of intelligence," "displays ambition," and "seeks reassurance from partner" (reversed).[6]

A second look at Tables 2 through 7 will reveal that the means and correlates of these components are reported amid the results derived from single items, discussed previously in this section. In general, the results derived from components are consistent with those from their associated

6. Full details of this principal components analysis are posted on our Web site.

items, with correlations and mean differences not dramatically different from those obtained at the item level.

We would not claim that this three-factor solution is the single underlying structure of the RBQ in all contexts or for all purposes. It might well differ across different situational contexts or different samples, and only further research will tell. We present these components only to illustrate the kind of more general and abstract-level information that might be obtained, for some purposes, by combining items with related content.

DISCUSSION

This article presents a tool for the description of social behavior. The Riverside Behavioral Q-sort has been used to describe experimental situations recorded on videotape, and psychometric analyses have shown that when four coders are employed the resulting reliabilities vary widely but are in general adequate. The most important testimony to the reliability of the RBQ comes from its demonstrated validity. Many RBQ items changed both significantly and meaningfully across experimental conditions (the unstructured and competitive situations) in a way that reflected their psychological dynamics. And in a parallel fashion, many RBQ items correlated both significantly and meaningfully with two different attributes of personality, extraversion and depression (BDI scores). It is this validity-in-use that provides the most powerful evidence that the RBQ has the potential to serve as a useful tool for personality and social psychology.

A few remaining issues are worthy of brief discussion.

Correlational and experimental use of the RBQ

It may be noted that although research with the RBQ to date has primarily employed correlational designs, experimental designs are possible as well (e.g., Funder & Colvin, 1991; Ozer, Funder, & Hershey, 1996). Analyses of situational effects, such as those presented in Funder and Colvin (1991), and in the Tables 2 and 3 in this article, demonstrate a large of number of behavioral differences that were a function of the situation.

When the nature of the behavioral context is experimentally manipulated, the RBQ offers the potential to assess the effects of a situational

independent variable on 64 behavioral dependent variables, rather than on just one, which is more typical. Thus, the potential usefulness of the RBQ is at least as great for experimental social psychology as for correlational studies within personality psychology.

Different items for different contexts?

As Bakeman and Gottman (1980) state, using a coding scheme developed by another researcher is "a little like wearing someone else's underwear" (p. 19). They go on to suggest that to avoid such an uncomfortable state of affairs, the researcher who needs a coding scheme should consider creating a new one or modifying an existing one to suit his or her particular needs. The RBQ was designed to provide information relevant to a wide variety of social interactions, but surely not every item would be relevant to every situation. For example, the item "Expresses awareness of being on camera or in an experiment" might be uninformative in naturalistic observation studies or studies in which a video camera was effectively concealed. We thus recommend that individuals using the RBQ as a coding scheme carefully consider each of the items and their relevance to the issues at hand. Perhaps some items should be excluded and other items used instead.

We would point out, however, that the items in the RBQ were not developed casually or quickly. They are the product of years of experience and continual refinement. Writing behavioral items that work, like writing test items in general, is more difficult than it may appear. So the possibility should also be considered that the RBQ could provide a standard basis for studies focused on interpersonal interactions. The wide-ranging content of the RBQ may recommend it as a basis for a common language through which to measure and discuss interpersonal behavior across different studies.

Q-sorts versus ratings

One significant feature of the RBQ in its typical format is the Q-Sort procedure itself, which has several advantages over typical rating methods. Not only are between-judge differences in mean ratings and standard deviations eliminated and some response sets minimized, but the judge is required to make fully considered ratings of each item against each other item. Consequently, ratings obtained through Q-Sort procedures

may be superior to ratings obtained through other methodologies. The Q-Sorting procedure, however, requires much more time than does the typical Likert-scale type of rating method, and the costs and benefits of sorting versus other means of rating items deserves attention in future research.

Back to behavior

Finally, we encourage researchers in personality and social psychology to renew their attention to behavior, one of the fundamental concerns of psychology. There are many understandable reasons why researchers so often avoid comprehensive assessment of the behavior of their research subjects. It is easier and less expensive to present several personality inventories to a group of participants than to obtain and analyze many kinds of behavioral data. In addition, the systematic observation and recording of behavior is still surprisingly rare in personality and social psychology, and thus there are few examples to serve as guides, or "off-the-shelf" techniques ready to be used.

Both comparative and developmental psychology have a long history of developing techniques for coding the behavior of their research subjects, perhaps because neither animals nor small children are cooperative about completing questionnaires. But personality and social psychology have often been lured by methodological ease to focus their attention away from the behavior of their human, adult subjects. As a result, we know much less than we should about the behaviors people perform in different social contexts, and the connections between behavior, situations, and personality. It is not our claim that the RBQ comprehensively samples observed personality, nor that it fully represents all varieties of social behavior. It is our hope that the development of the RBQ provides a decent start toward these ultimate goals, and that further research will continue and extend the investigation and classification of observed social behavior.

REFERENCES

Alden, L. E., & Phillips, N. (1990). An interpersonal analysis of social anxiety and depression. *Cognitive Therapy and Research*, **14**, 499–513.

Atkinson, R. L., Atkinson, R. C., Smith, E. E., & Bem, D. J. (1993). *Introduction to psychology* (11th ed.). Ft Worth, TX: Harcourt Brace Jovanovich College Publishers.

Bakeman, R., & Gottman J. M. (1980). *Observing interaction: An introduction to sequential analysis*. Cambridge, UK: Cambridge University Press.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, **4**, 561–571.

Bem, D. J., & Funder, D. C. (1978). Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review*, **85**, 485–501.

Block, J. (1977). Advancing the psychology of personality: Paradigmatic shift or improving the quality of research. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology* (pp. 37–64). Hillsdale, NJ: Lawrence Erlbaum Associates.

Block, J. (1978). *The Q-sort method in personality assessment and psychiatric research*. Palo Alto, CA: Consulting Psychologists Press. (Originally published 1961).

Cairns, R. B., & Green, J. A. (1979). How to assess personality and social patterns: Observations or ratings? In R. B. Cairns (Ed.), *The analysis of social interactions: Methods, issues and illustrations* (pp. 209–226). Hillsdale, NJ: Erlbaum.

Caspi, A., Block, J., Block, J. H., Klopp, B., Lynam, D., Moffitt, T. E., & Strouthamer-Loeber, M. (1992). A "common language" version of the California Child Q-Set for personality assessment. *Psychological Assessment*, **4**, 512–523.

Colvin, C. R., Block, J., & Funder, D. C. (1995). Overly positive self-evaluations and personality: Negative implications for mental health. *Journal of Personality and Social Psychology*, **68**, 1152–1162.

Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.

Creed, A. T., & Funder, D. C. (1997). Social anxiety: From the inside and outside. *Personality and Individual Differences*, **25**, 19–33.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin*, **52**, 177–193.

Eaton, L. G., & Funder, D. C. (1999). *The creation and consequences of the social world: An interactional analysis of extraversion*. Manuscript under review.

Ekman, P. W., & Friesen, W. (1978). *Manual for the facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.

Ellgring, H. (1989). *Nonverbal communication in depression*. Cambridge, UK: Cambridge University Press.

Funabiki, D., Bologna, N. C., Pepping, M., & Fitzgerald, K. C. (1980). Revisiting sex differences in the expression of depression. *Journal of Abnormal Psychology*, **89**, 194–202.

Funder, D. C. (1991). Global traits: A neo-Allportian approach to personality. *Psychological Science*, **2**, 31–39.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, **102**, 652–670.

Funder, D.C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego: Academic Press.

Funder, D. C., Block, J. H., & Block, J. (1983). Delay of gratification: Some longitudinal personality correlates. *Journal of Personality and Social Psychology*, **44**, 1198–1213.

Funder, D. C., & Colvin, C. R. (1991). Explorations in behavioral consistency: Properties of persons, situations, and behaviors. *Journal of Personality and Social Psychology*, **60**, 773–794.

Furr, R. M., & Funder, D. C. (1998). A multi-modal analysis of personal negativity. *Journal of Personality and Social Psychology*, **74**, 1580–1591.

Gjerde, P. F., Block, J., & Block, H. H. (1988). Depressive symptoms and personality during late adolescence: Gender differences in the externalization-internalization of symptom expression. *Journal of Abnormal Psychology*, **97**, 475–486.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, **48**, 26–34.

Hammen, C. L., & Padesky, C. (1977). Sex differences in the expression of depressive responses on the Beck Depression Inventory. *Journal of Abnormal Psychology*, **86**, 609–614.

Hammond, K. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. New York: Oxford University Press.

Hokanson, J. E., Lowenstein, D. A., Hedeen, C., & Howes, M. J. (1986). Dysphoric college students and roommates: A study of social behaviors over a three-month period. *Personality and Social Psychology Bulletin*, **12**, 311–324.

Horowitz, L. M. (1979). On the cognitive structure of interpersonal problems treated in psychotherapy. *Journal of Consulting and Clinical Psychology*, **47**, 5–15.

Kahneman, D. T., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, **80**, 237–251.

Kalbaugh, P. E., & Haviland, J. M. (1994). Nonverbal communication between parents and adolescents: A study of approach and avoidance behaviors. *Journal of Nonverbal Behavior*, **18**, 91–113.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, **43**, 23–34.

LaForge, R., & Suzek, R. (1955). The interpersonal dimensions of personality: III. Interpersonal checklist. *Journal of Personality*, **24**, 94–112.

McCrae, R.R., Costa, P.T., Jr., & Busch, C.M. (1986). Evaluating comprehensiveness in personality systems: The California Q-set and the five factor model. *Journal of Personality*, **54**, 430–446.

Mischel, W. (1968). *Personality and assessment*. New York: John Wiley and Sons.

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, **80**, 252–283.

Montepare, J. M., Goldstein, S. B., & Clausen, A. (1987). The identification of emotions from gait information. *Journal of Nonverbal Behavior*, **11**, 33–42.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.

Nolen-Hoeksema, S. (1987). Sex differences in unipolar depression: Evidence and theory. *Psychological Bulletin*, **101**, 259–282.

Nolen-Hoeksema, S. (1990). *Sex differences in depression*. Stanford, CA: Stanford University Press.

Ozer, D. J. (1993). The Q-sort method and the study of personality development. In D. C. Funder, R. D. Parke, C. Tomlinson-Keasey, & K. Widaman (Eds.),

Studying lives through time: Personality and development (pp. 147–168). Washington, DC: American Psychological Association.

Ozer, D. J., Funder, D. C., & Hershey, J. W. (1996, November). *Behavior as a function of the person and the situation: Generalizability analyses*. Paper presented at the Annual Meetings of the Society for Multivariate Experimental Psychology, Evanston, IL.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, **86**, 420–428.

Stephenson, W. (1953). *The study of behavior: Q-technique and its methodology*. Chicago: University of Chicago Press.

Vredenburg, K., Krames, L., & Flett, G.L. (1986). Sex differences in the clinical expression of depression. *Sex Roles*, **14**, 37–49.

Wiehl, L.G. (1997). *Predicting behavior at the crossroads of social and personality psychology: The creation and consequences of the social world*. Unpublished doctoral dissertation, University of California, Riverside.

Youngren, M. A., & Lewinsohn, P. M. (1980). The functional relation between depression and problematic interpersonal behavior. *Journal of Abnormal Psychology*, **89**, 333–341.

Zimbardo, P. G., & Weber, A. L. (1994). *Psychology*. New York: HarperCollins.

Appendix A

Behavioral Q-Sort

Coding Procedures:

Coders were recruited from upper-division psychology courses and were given course credit for their efforts. The coders were trained and were provided with the following set of instructions:

Instructions

Do the Q-sort in two phases. First, watch the entire videotape. Then, divide the Q-items into three piles. Many of the items will not be particularly relevant to characterizing the subject's behavior. These items, which should include roughly half of the total set, should be placed into a middle pile. The items that do characterize the subject's behavior go in the right-hand pile, and those that negatively characterize the subject's behavior (i.e., characterize his or her behavior in the opposite direction to the way the item is worded) go in the left-hand pile.

Then, watch the tape again, paying particular attention to any questions or uncertainties that might have arisen during the initial division of the

cards. Then do a complete Q-sort. The final distribution is into nine categories ranging from negatively characteristic (1) to highly characteristic (9). Items that do not seem relevant should go into the middle category (5). The number of cards to be placed in each category is 3, 5, 7, 10, 14, 10, 7, 5, 3.

Two important notes about how to sort items:

(1) We are trying to characterize fairly directly the *behavior that can be seen on the videotape.* So, as much as possible, avoid drawing inferences about characteristics or intentions of the subjects that are not directly visible. Direct statements of feelings such as guilt, insecurity, or confidence should be taken at face value, and such feelings should not be inferred in the absence of direct statements or other visible evidence.

(2) Items should be placed according to their importance in distinguishing this subject's behavior from the behavior of other subjects on the other tapes of the same situation. Thus, items should receive very high (or low) placement primarily to the degree that they serve to distinguish or set apart this subject's behavior from that of the other subjects. This means that some experience with the range of behavior seen is necessary before one can begin Q-sorting any particular situation.