The Effects of Biasing Information on Behavioral Observations and Rating Scales

Lawrence J. Siegel
University of Missouri Medical School

Susan L. Dragovich
Case Western Reserve University

David Marholin II
Boston University

The effects of biasing information on behavioral observations and rating scales were studied. Forty-one undergraduate students trained in making reliable behavioral observations were given differential expectations concerning the activity level of a target child. They then viewed videotape recordings of that child and tallied frequency counts of six behavioral categories simultaneously. In addition, subjects completed postexperimental rating scales composed of specific, identifiable behaviors in regard to the target child. Results indicated that, for the most part, neither the behavioral observations nor the rating scales were significantly affected by the biasing information. It is suggested that rating scales constructed of items as discrete and readily identifiable as those of behavioral observation measures may prove resistant to biasing effects.

With a recent increase in the development and use of behavioral techniques for the treatment of behavior disorders (Bijou & Redd, 1975), there has been a
growing reliance on behavioral observation measures to assess both the behavior problem and the effectiveness of the chosen treatment program (Kazdin, 1975; Marholin & Bijou, in press). Indeed, in many studies behavioral observation is the only method which is employed to secure data. Given the importance of accurate assessment for treatment program changes, it is crucial to provide reliable behavioral observation procedures (Baer, Wolf, & Risley, 1968). Several investigators, however, have questioned the consistency of reliability under numerous conditions (Morris & Rosen, in press) such as covert versus overt reliability procedures (Reid, 1970), experimenter instructions and feedback to observers (O’Leary & Kent, in press; O’Leary, Kent, & Kanowitz, 1975; Taplin & Reid, 1973), prior information and behavioral predictability (Mash & Makohoniuk, 1975), consensual observer drift (O’Leary & Kent, 1973), and observer training and the complexity of coding categories (Mash & McElwee, 1974; Wildman, Erickson, & Kent, 1975).

One factor affecting the validity of behavioral observations which has received little attention by researchers is that of observer bias. Rosenthal and Jacobson’s Pygmalion in the Classroom (1968), though criticized on methodological grounds (e.g., Thorndike, 1969), has served to focus attention on how intentionally introduced biasing information given teachers may affect their observations and subsequent IQ ratings of children in the direction of the biasing information. Although these particular findings have not been unequivocally replicated, the implied threat of observer bias to the validity of behavioral observations nevertheless deserves further investigation.

One area of the literature which has gathered much of its data via behavioral ratings and direct observational procedures, and which has been described as subject to the effects of biasing information, is that of drug treatment with hyperactive children (Werry & Sprague, 1970). Most studies assessing the effects of drugs on the behavior of hyperactive children utilize some form of behavioral observation technique or rating scale in comparing the pre- and postdrug behavior of the children (e.g., Conners, 1969, 1970; Conners & Eisenberg, 1963; Conners, Eisenberg, & Barcia, 1967; Eisenberg, 1966; Quay, Sprague, Schulman, & Miller, 1966). Sulzbacher (1973) and Sroufe (1973) have criticized the use of rating scales in drug research with hyperactive children, suggesting that the validity of such scales is significantly affected by various kinds of information. Rating scales typically employ globally descriptive terms (e.g., inattentive, disruptive, aggressive, or improved vs. unimproved) that call for inferential decisions on the part of raters. Therefore, it is difficult to assess from subsequent ratings which particular behaviors raters were actually attending to in arriving at their impressions. In addition, since the rating scale items often have not been anchored to concrete behaviors, the raters’ global inferences are more susceptible to biasing information introduced into the setting. For example Sulzbacher (1973) reported that rating scales completed by nurses regarding hyperactive children receiving either drug or placebo significantly favored the children on the