Methodological Issues in Interviewing and Using Self-Report Questionnaires With People With Mental Retardation

W. M. L. Finlay and E. Lyons
University of Surrey

In this article the authors review methodological issues that arise when interviews and self-report questionnaires are used with people with mental retardation and offer suggestions for overcoming some of the difficulties described. Examples are drawn from studies that use qualitative methodology, quantitative studies assessing different question types, and studies reporting on the development of instruments measuring psychiatric symptoms, self-concept, and quality of life. Specific problems that arise with respect to item content (e.g., quantitative judgments, generalizations), question phrasing (e.g., modifiers), response format (e.g., acquiescence, multiple-choice questions), and psychometric properties (factor structure and validity) are discussed. It is argued that because many self-report questionnaires include questions that have been found to be problematic in this population, more attention needs to be paid to establishing the validity of such measures and to clearly defining the population for which the instrument is designed.

The importance of obtaining information from people with mental retardation is increasingly stressed by researchers, clinicians, and those involved in service development. This is a result of the limitations of informant reports for many psychological phenomena (such as subjective states), as well as a change in the philosophy of services toward user consultation and a greater emphasis on people with mental retardation making decisions about their own lives. Structured or semistructured interviews are used in the formulation stage of many psychological therapies, in the assessment of abilities, IQ, and neuropsychology, as well as in forensic, research, and individual planning contexts. Verbal self-report questionnaires are used in clinical assessment (e.g., to assess psychopathology, self-esteem, personality), research, and service evaluation. However, difficulties in conducting interviews and using self-report questionnaires with this population are widely reported. Such difficulties have been described in psychiatric and psychological assessments (e.g., Coelho & Saunders, 1996; Duckworth, Radhakrishnan, Nolan, & Fraser, 1993; Fraser, Leudar, Gray, & Campbell, 1986; Rojahn, Warren, & Ohringer, 1994; Scott, 1994; Sovner & Hurley, 1986; Zetlin, Heriot, & Turner, 1985), police interviews (Clare & Gudjonsson, 1995; Tully & Cahill, 1984), service evaluations (e.g., Antaki & Rapley, 1996; Dagnan, Dennis, & Wood, 1994; Flynn, 1986; McVilly, 1995; Smyley & Ellsworth, 1997), vocational assessment (Wilgosh & Barry, 1983), and the assessment of service needs and choices (e.g., Gow & Balla, 1994; Lindsey, 1994).

The problems of communication associated with mental retardation are well documented (e.g., Abbeduto & Hagerman, 1997; Bedrosian, 1993; Beveridge, Conti-Ramsden, & Leudar, 1997; Brinton & Fujiki, 1993, 1994; McLean, Brady, & McLean, 1996; Sabsay & Kerman, 1983). These include difficulties in the production of symbols, lack of understanding of complex grammatical structures or concepts, and difficulties in designing utterances for the specific, current interactional context (e.g., the inclusion of appropriate content). As a result of these difficulties, interviewers must pay particular attention to providing statements or questions to which the person with mental retardation can respond and to addressing problems in understanding or classifying the responses given.

In many cases, the use of questionnaires developed for the general population is inappropriate for people with mental retardation because of the respondents' inability to comprehend the question and express an answer clearly and because the psychometric properties may not be applicable to this population. For this reason, many new self-report questionnaires for adults with mental retardation have been developed assessing psychiatric symptoms (e.g., Kazdin, Matson, & Sensore, 1983; Lindsay & Michie, 1988; Moss et al., 1993; Reynolds & Baker, 1988; Stroghmer, Prout, & Gorsky, 1994), self-concept (e.g., Szivos-Bach, 1993), quality of life (e.g., Barlow & Kirby, 1991; Heal & Chadsey-Rusch, 1985; Schalock, Keith, Hoffman, & Karon, 1989), as well as a range of other constructs. These include community living (Botwick & Foss, 1981), loneliness (Chadsey-Rusch, DeStefano, O'Reilly, Gonzalez, & Collet-Klingenberg, 1992), knowledge of rights.
(Flynn, Reeves, Whelan, & Speake, 1985), individual choice (Stancliffe, 1995), stigma, and social support (Reiss & Benson, 1985).

Despite the widespread use of self-report instruments with this population, however, the validity of many questionnaires has not been established. There are a number of difficulties associated with asking questions of people with mental retardation that should lead to particular concerns about validity. The following sections describe problems with question content, phrasing, response format, and the psychometric properties of self-report questionnaires. Because self-report questionnaires are usually administered orally rather than in written form in this population, issues described below concerning the use of interviews in general are also applicable to the design and use of questionnaires. Although people with mental retardation are a heterogeneous population, there are, unfortunately, no adequate data to specify the subgroups for which each problem is particularly important. Although techniques for overcoming problems that have been suggested in the literature are described wherever possible, in many cases there are no ready solutions, and more research is needed into strategies to improve questioning in this population.

Question Content

It is widely acknowledged that vocabulary and meaning should be clear and simple (e.g., Lowe & de Paiva, 1988; Prosser & Bromley, 1998). The person may be unfamiliar with the concepts and vocabulary that the interviewer uses or may be unable or unaccustomed to making the type of judgments requested. The following sections describe difficulties that arise when questions concern quantitative judgments, comparisons, abstract concepts, inferences about the attitudes of others to the self, generalizations, and unfamiliar or sensitive content. Problems of content can be found in questionnaires assessing mental states, frequencies of events and symptoms, social judgments, and abstract concepts such as autonomy.

Quantitative Judgments and Time Questions

Questions that are about time or that require a judgment of frequency or degree have been found to be problematic for many people with mental retardation (Biklen & Moseley, 1988; Booth & Booth, 1994a, 1994b, 1996; Flynn, 1986; Lindsay & Michie, 1988; Malik, Ashton-Shaeffer, & Kleiber, 1991; Matson & Frame, 1986; Moss et al., 1997; Wyngaarden, 1981). For example, Sigelman, Schoenrock, et al. (1981) found that in samples of institutionalized and community-living adults and children with severe to mild mental retardation, multiple-choice questions requiring a quantitative judgment produced low levels of agreement with observers and with alternative question formats. This calls into question the use of Likert-type questions to assess degree or frequency, such as in the modified versions of the Zung Self-Rating Depression Scale (ZDS) and the Beck Depression Inventory (BDI; Helsel & Matson, 1988; Kazdin et al., 1983; Matson, Kazdin, & Senatore, 1984; Senatore, Matson, & Kazdin, 1985). The modified ZDS, for example, requires the respondent to make a frequency estimate for certain emotional states, choosing among the following: none or a little of the time, some of the time, a good part of the time, and most of the time. A visual aid (bar graph) is often given to facilitate such judgments, although the efficacy of this method remains to be demonstrated. Although it may enable the range of options to be remembered more easily, it is unlikely to overcome the difficulties that people may have with making the estimate itself.

Screening questions concerning concrete activities whose frequency has been established from informants might be used in pretest sessions to ascertain the ability of individuals to make such judgments. In some cases, it may be necessary to use simple yes-no questions instead. Lindsay and Michie (1988), in a sample of 29 people with mild and moderate mental retardation, found yes-no questions to produce greater split-half reliability (r = .69) than 4-point estimates of frequency (r = .12) in the assessment of anxiety. Because of the importance of frequency estimates in the assessment of many characteristics, further research needs to be carried out to determine the effectiveness of visual aids and to develop other methods for eliciting such estimates.

To facilitate estimations of time, significant events in the participants’ lives have been used. In a longitudinal, interview-based study of the experiences of 20 parents who had mental retardation, Booth and Booth (1994b) used situational markers rather than asking for specific dates or time periods. When the dates of certain events were known (e.g., a child’s age, the place the person was living at the time), the timing of other events could be determined with respect to them. Similarly, the self-report Psychiatric Assessment Schedule for Adults with Developmental Disability (PAS-ADD; Moss et al., 1993; Patel, Goldberg, & Moss, 1993) requires the interviewer to identify (from informants) a significant event in the person’s life that has occurred about 1 month before the interview. These events are used in the self-report interview to assess the respondent’s concept of time and as a marker to gauge the recency of any symptoms.

Direct Comparisons

Questions concerning direct comparisons have been found to be difficult (Biklen & Moseley, 1988; Heil & Sigelman, 1995; Matson & Frame, 1986; Smyley & Ellsworth, 1997). Comparisons are often requested in interviews assessing changes in people’s lives, their preferences, symptom severity (e.g., Matson et al., 1984), and in social comparison questionnaires themselves (e.g., Dagnan & Sandhu, 1999). For example, the modified ZDS contains a range of items asking the respondent to compare their current emotions or behavior with the past (e.g., “You can think as good as you used to think”); Helsel & Matson, 1988).

If comparisons are difficult for respondents, it may be better to ask about one element at a time (Biklen & Moseley, 1988; Catermole, Jahoda, & Markova, 1990; Heil & Sigelman, 1985). For example, Smyley and Ellsworth (1997), in their study of satisfaction with day services, asked first about the participants’ old day center and then about the new one. The results were content analyzed and compared to infer positive, negative, or neutral comparisons. When individuals are able to make direct comparisons and there are a number of items that are to be ranked, dichotomous choices can be given, in which each item is paired with every other item (e.g., Botswick & Foss, 1981; Dagnan et al., 1994). However, respondents may become bored and may question why they are responding repeatedly to similar questions.
Socially Reflexive Questions

Questions asking people how they think others view or evaluate them are difficult for many people (Szivos-Bach, 1993). Such questions involve a complex level of social understanding because they require the informant to infer the internal emotions, attitudes, or beliefs of another person from their behavior. Informants may be unaccustomed or unable to make such judgments. Socially reflexive items are often found in questionnaires assessing self-esteem and depression (e.g., Matson et al., 1984; Reynolds & Baker, 1988) and perceptions of social situations (e.g., stigma—Reiss & Benson, 1985; quality of life—Schalock et al., 1989). Examples from the Psychopathology Instrument for Mentally Retarded Adults (PIMRA—Helsel & Matson, 1988; Kazdin et al., 1983) include, “Do people have trouble understanding what you say?”, “Do people know how you feel?”, and “Do people like to be with you?” Simple yes–no answers to such questions may disguise a guess or an incorrect understanding of the question.

Abstract Concepts and Generalized Judgments

Researchers often recommend the use of questions that refer to specific activities or events rather than those referring to abstract concepts because these are more easily understood and evaluated (e.g., Booth & Booth, 1994; Smyley & Elsworth, 1997). Problems may result from difficulties in aggregating specific instances to make a general evaluation, from a lack of expressive abilities, or from a lack of understanding of certain concepts. For example, in a sample of 45 defendants with mental retardation (mean IQ = 61, SD = 6.18), Smith (1993) found that 7 were unable to understand the term guilty, and 9 did not understand not guilty.

Questions about emotions have been found to be harder to answer than those about concrete situations for some people with mental retardation (Booth & Booth, 1994b; Lowe & de Paiva, 1988; Malik et al., 1991; McVilly, 1995; Sigelman, Winer, & Shearn, 1997). Any questions using official labels for this population were administered to 46 adults with mild to severe mental retardation (mean IQ = 51, SD = 10) employed in a sheltered workshop, Zetlin et al. (1985) found that most of the 35% of initial answers judged to be problematic took the form of specific, single examples of when the event occurred or the person felt that way. It was often difficult to establish whether such experiences were frequent occurrences—the answers never reached the level of generalization required by the question. Antaki (1998), using extracts from quality-of-life interviews, demonstrated how interviewers anticipated this difficulty and, in their efforts to paraphrase, restricted the meaning of general questions to specific occasions or activities. Antaki found that word-for-word delivery of questions as written was rare and that such paraphrasing was the norm.

Concepts are also problematic when the respondent interprets the terms with either a more general or a more restricted and context-specific definition than the researcher. This is a problem with any population but may be particularly acute in populations in which individuals’ life situations are very different from that of the interviewer. In these cases, even terms that seem to have clear referents may be interpreted differently. For example, Barlow and Kirby (1991), in a study of the life satisfaction of 31 adults with mild mental retardation living either in the community or in an institution, found that participants frequently interpreted the term friends in a more general way than the researchers, including acquaintances who acted in a friendly manner to them.

Differences in reference might also occur because of differences in the comparative context in which a question is framed. Stancliffe (1995) carried out a study with 46 people with severe to borderline mental retardation who were clients of supported accommodation agencies in which the amount of choice that the participants reported over various aspects of their lives was compared with the amount of choice that the staff reported. Reverse wordings were used (“Do you choose . . . ?” and “Does someone else choose . . . ?”) with 4-point Likert scales. After contradictions were removed, it was found that clients’ self-ratings were significantly greater (i.e., more choice) than staff self-ratings for 3 of the 10 items. These concerned spending (Ms = 3.79 and 3.48, respectively), living companions (Ms = 3.40 and 2.60, respectively), and employment (Ms = 3.30 and 2.33, respectively). Stancliffe explained this as being due to the two groups using a different set of social comparison targets and due to the clients having a limited experience of the full range of possible choices.

Questions that are prone to such difficulties are found in questionnaires that assess internal states and emotions and in those assessing concepts such as autonomy and responsibility (e.g., Lindsay, 1994; Reiss & Benson, 1985; Schalock et al., 1989). For example, in a study of self-determination, Wehmeyer and Metzler (1995) asked questions such as, “Do you determine which agencies and organisations provide services and support for you?” and, “Do you usually choose your friends and acquaintances?” Further questions asked participants how integrated, independent, and productive they were. The extent to which abstract or nonspecific questions are problematic depends on the aims of the interview. If the interviewer has specific definitions and standards in mind, then such questions should be avoided or at least followed by further (scripted) questions to establish the meaning for the participant. If, however, the interviewer is more concerned with the meanings of the respondent, then such questions are appropriate, although follow-up questions are also necessary in these cases.

Unfamiliar Content

Through the content of questions or the response options provided, questionnaires might require respondents to consider themselves or their lives in uncharacteristic ways. This is a problem of content validity and may be found in questionnaires assessing self-concept, psychiatric symptoms, and quality of life. Schurr, Joiner, and Towne (1970) pointed out that instruments used to assess self-concept may often reflect the concerns of the researcher rather than the respondent (see also Gowans & Hulbert, 1983; Nooe, 1977; Rapley, 1995; Zetlin et al., 1985). Relevant content is also an issue in psychiatric assessment, and some researchers have questioned whether psychiatric disorders exhibit the same symptoms in the general population as they do when combined with mental retardation (e.g., Aman, 1991; Moss et al., 1993; Moss, Prosser, & Goldberg, 1996; Sovner & Hurley, 1986). Caution should therefore be exercised in modifying questionnaires developed for the general population by simplifying the language but otherwise leaving the item content intact (e.g., Jiranek & Kirby, 1990; Kazdin et al., 1983; Lindsay & Michie, 1988; Lindsay, Michie, Baty, Smith, & Miller, 1994; Prout & Schaeffer, 1985; Reynolds & Miller, 1985).

Concern has been expressed in the use of quality-of-life questionnaires that select a number of criteria judged by professionals to be important indicators of service quality, such as adaptive behaviors or aspects of the physical environment (Bercovici, 1981; Heal & Chadsen-Rusch, 1985; Taylor & Bogdan, 1981). When judging the success of community services, Cattermole et al. (1990) suggested that it is more important to assess quality of life through finding out what the service users value rather than what professionals assume to be important. Using an open-ended interview style and frequency counts, they found in a sample of 15 people who had moved to staffed homes in the community that personal autonomy and social life were of more concern to participants than were learning skills (see also Dudley, Calhoun, Ahlgrim-Dezelle, & Conroy, 1997).

A range of procedures suggested in the general literature on psychological assessment can be used to increase the content validity of new measures (e.g., Haynes, Richard, & Kubby, 1995). After initial items are generated, they should be reviewed by multiple judges to ensure that they sample all possible facets of the construct of interest and that they do not include items that measure correlates rather than the construct itself (Clark & Watson, 1995; Haynes et al., 1995; for an example, see Thorin, Brownig, & Irvin, 1988). Initial items, as well as instructions, can be generated through expert sampling and through pilot work with people with mental retardation (e.g., Botswick & Foss, 1981; Smyley & Ellsworth, 1997). For example, Castles and Glass (1986) developed a social self-efficacy questionnaire based on problem vignettes that were generated from interviews with 39 people with mental retardation in a vocational training facility and 42 professionals. Similarly, Noe (1977) collected self-relevant statements from a small sample of people with mental retardation living in a residential facility and used them as a basis for a card-sorting task to assess self-concept. In such cases, it is important that the sample characteristics are well described because content developed from people with mild mental retardation, or from those living in institutions, may not be relevant for those with moderate disabilities or those living in the community.

Sensitive Content

Questions with sensitive or taboo content have been found to be prone to error and bias in the general population (for a review, see Barnett, 1998). These problems may be accentuated in people with mental retardation, who often use services in which professionals exert a large degree of control over their lives. They may, therefore, be concerned with the possible consequences of their responses, particularly as service use is characterized by the sharing of information among professionals (Biklen & Moseley, 1988; Prosser & Bromley, 1998). Because many people do not have sufficient reading abilities, self-report research with this population tends to involve face-to-face interviews, which are less anonymous and private than written questionnaires and which may lead to underreporting of certain behaviors. Pack, Wallander, and Browne (1998), in a study involving 82 adolescents with mild mental retardation who were in special education programs, found that significantly more participants reported engagement in risky behaviors (e.g., drug use, carrying weapons) when the group administration of an anonymous survey was used compared with a confidential interview (Ms = 23.8% and 13%, respectively). In addition, sensitive questions may lead to different errors or biases than in the general population because responding in a socially desirable way requires attention to the particular form of the question. Shaw and Budd (1982) carried out a study in which 24 members of a sheltered workshop (mean IQ = 49.3, SD = 13.2) were asked two reverse-worded questions about a list of behaviors (“Is it against the rules to . . . ?” and, “Are you allowed to . . . ?”). They found that participants were more likely to say “no” to questions concerning prohibited behavior compared with desirable behavior, regardless of the form of the question (mean nay-saying questions = 13.7 and 2.9, respectively).

Barnett (1998) described a variety of techniques that can be used to reduce the impact of sensitive questions. Those that may be useful with people with mental retardation include explicit statements that information will not be shared with caregivers or service workers (see Prosser & Bromley, 1998, for further suggestions on explaining research to participants), use of vignettes or nominative techniques (e.g., asking what other people would do), and the use of open-ended questions and conversational styles.
Question Phrasing

The use of complex sentence structures in questions creates difficulties for many people with mental retardation (Kabzems, 1985; McConkey, Morris, & Purcell, 1999; Prosser & Bromley, 1998; Wyngaarden, 1981). The following sections describe problems that arise when questions contain negatives and other modifiers, passive tenses, examples, and instances in which the subject and object can be confused. Many problems with questions and answers may be due to participants responding to particular words in the question rather than to the whole (Matson & Frame, 1986; Zetlin et al., 1985).

Negatively Worded Questions and Modifiers

It has been shown in experimental studies using reaction times and error rates that negatively phrased items (i.e., when no or not are added to positive phrases) are more difficult to respond to than are affirmative items for the general population (Gough, 1965; Slobin, 1966). Negatively worded questions have been reported as difficult for people with mental retardation (Lowe & de Paiva, 1988; Welnhmers, 1994) and are an example of the more general problem of modifiers, which are single words or clauses that change the sense of a question. In these cases, the person may respond to the simple form of the item as if the modifier were not present. For example, “What things would you like to change about yourself?” might be responded to as if it were, “What things would you like to change?”

Not only are negatively phrased questions more complicated constructions, leading some people with mental retardation to answer as if the question had been phrased in the positive, but members of this population might be less willing to criticize than to offer positive comments (Lowe & de Paiva, 1988). Negatively phrased questions can be found in questionnaires assessing affect, self-esteem, and preferences, particularly when there is no commonly used single word for the negative form of a concept. They are often included to counterbalance items phrased positively. Examples include the 1985 assessment by Flynn et al. of people’s tolerance of rules, the Eysenck-Withers Personality Inventory for Subnormal Subjects (Eysenck, 1965), and the Self-Report Depression Questionnaire (Reynolds & Baker, 1988). Items in the latter include, “I can’t fall asleep at night,” “I have no energy,” and, “I feel that people don’t like me.”

Because of these difficulties, it is better to use affirmative forms of statements. For example, Szivos-Bach (1993), basing her questionnaire on standard self-esteem questionnaires, used statements such as, “I cause trouble/in give up easily/I am slow at work/I make a mess of things I try/I forget things.” Statements such as these are better than simply adding a negative to a positive (e.g., “I am not very good at . . . /I can’t remember things”). However, it is important to ensure that the words used are common in the person’s everyday environments (e.g., terms such as dislike or unkind might not be used).

Subject–Object Confusion and Passive Phrasing

Slobin (1966) found, for the general population, that sentences containing the passive tense were more difficult than those using the active tense (see also Gough, 1965) and that reversible items, in which the subject and object can be confused although the sentence still makes sense (e.g., the man chases the dog), were harder than nonreversible items (e.g., the man waters the flowers). This may be a particular problem for people with mental retardation, who might confuse the subject and object of the question, particularly if they attend to only a few words per question.

Flynn et al. (1985), in a questionnaire to assess people’s tolerance of rules and knowledge of rights, included questions prone to subject–object difficulties such as, “Do you think you should be careful with things that do not belong to you?” and, “Do you think you should blame other people when you do something wrong?” Questionnaires that attempt to assess people’s relationships with, and reactions to, other people are also prone to such difficulties (e.g., Eysenck, 1965; Matson et al., 1984).

Giving Examples

In a study in which 107 children and 42 adults with mild to severe mental retardation living in both the community and institutions were interviewed about their activities, quality of life, and residential settings, Sigelman, Budd, et al. (1982) found that when they gave examples, people tended to give these back as answers. This particularly may be a problem when respondents have difficulties in comprehension, memory retrieval, making generalizations, or with perseveration.

Response Format and Scoring Answers

Self-report questionnaires often involve yes–no or multiple-choice response formats. For example, the most widely used instruments that assess self-esteem, psychiatric symptoms, and quality-of-life and service satisfaction use yes–no, either–or, or multiple-choice or Likert scale formats. The following sections address issues arising from the use of yes–no, multiple-choice, and open-ended questions, as well as the problem of how to deal with irrelevant answers.

Yes–No Questions and Acquiescence

Acquiescence in interviews, or yea-saying, is the tendency to say yes to questions regardless of their content (Block, 1965). It is the problem that has been described most often in the literature on response biases in this population (for a review, see Finlay & Lyons, in press) and is linked to research suggesting that people with mental retardation are more suggestible to leading questions (Kebbell & Hatton, 1999).

In a series of studies using alternative question formats with children and adults with severe to mild mental retardation, Sigelman and her colleagues (Sigelman, Budd, Spanhel, & Schoenrock, 1981a, 1981b; Sigelman, Budd, Winer, Schoenrock, & Martin, 1982; Sigelman, Schoenrock, et al., 1981; Sigelman, Winer, et al., 1982; for a review, see Heal & Sigelman, 1995) found that for factual as well as subjective content, yes–no questions were subject to a systematic acquiescence bias, even when the answer was absurd (e.g., “Does it usually snow in the summer here?”). Sigelman, Budd, et al. (1981b), for example, found in a sample of 29 people with moderate and severe mental retardation (mean IQ = 31.62, SD = 7.24) that acquiescence rates ranged from 20% to 83.3% (depending on the type of question and method
of detection). This tendency was inversely related to IQ scores on two of their three composite measures of acquiescence ($r = .37$ and .61; see also Burnett, 1989; Gadjonsson, 1990; Shaw & Budd, 1982; cf. Mattika & Vesala, 1997). Either—or questions were found to be less prone to systematic response bias, as measured by reverse wordings, although there was a tendency in these cases for last-choice responding (see also Loper & Reeve, 1983). For example, Sigelman, Budd, et al. (1981a) found that in a sample of 42 institutionalized adults (mean IQ = 39.8, SD = 13.1), consistency when using reverse wordings was greater for either—or questions (86.2% and 75.1%) than for yes–no questions (53.1% and 52%). Sigelman, Winer, et al. (1982), reporting on the same sample, also found responsiveness to be highest with yes–no questions (82.2%), followed by either–or questions (68.7%), and to be lowest with open-ended questions (51%). These papers stressed that the validity of answers given by people with mental retardation must never be assumed but must always be demonstrated. Acquiescent responding to yes–no questions has also been found to be a problem by other authors (Burnett, 1989; Gerjuoy & Winters, 1966; Heal & Chadsey-Rusch, 1985; Perlman, Ericson, Esses, & Isaacs, 1994).

A number of studies have suggested that acquiescence may not be as common in people with mental retardation as suggested by Sigelman and colleagues’ studies (e.g., Booth & Booth, 1994a; Conroy & Bradley, 1985; Matikka & Vesala, 1997; Rapley & Antaki, 1996; Wehmeyer, 1994). Differences in the rates of acquiescence found across studies may be due to differences in the methods used and the samples selected, which unfortunately make the results noncomparable. Researchers have tended to use three techniques to identify acquiescence in interviews, nonsense questions to which the answer should be “yes,” pairs of questions that are opposite in meaning, and pairs of questions that ask the same question in different formats. Each of these techniques involves difficulties in interpretation. Nonsense questions (e.g., Flynn et al., 1985; Sigelman, Budd, et al., 1981b) have been criticized by Rapley and Antaki (1996), who questioned whether answers to such questions are comparable to those from sensible questions. Such questions may baffle or amuse participants, who may say “yes” for reasons other than a tendency to acquiesce.

Another method used is to include pairs of reverse-worded questions. If a respondent affirms both versions, this is taken as evidence of acquiescence (e.g., Flynn et al., 1985; Heal & Chadsey-Rusch, 1985; Sigelman, Budd, et al., 1981a, 1981b). However, Matikka and Vesala (1997) suggested that contradictions may arise because different images or occasions are being elicited, suggesting that using reverse wordings might not be a good way to measure acquiescence. Indeed, Rapley and Antaki (1996) pointed out that inconsistency is a feature of discourse in general. Despite the problems, however, Sigelman, Budd, et al. (1981b) offered some evidence for the validity of these techniques. They found significant correlations between measures based on reverse wordings and nonsense questions ($r = .57$) and between those based on nonsense questions and total number of affirmative responses to questions about involvement in household chores ($r = .45$).

Yea-saying may be more likely to arise when the answer is not known or when questions are too long or the structure is too complex (Finlay & Lyons, in press). Respondents may then focus on a limited part of the question and miss the subtleties of the phrasing. In these cases, participants may appear to contradict themselves when questions are reverse worded. In the study described earlier, Shaw and Budd (1982) found significantly more acquiescence for socially desirable than undesirable behaviors ($Ms = 13.3$ and 4.6 questions, respectively), significantly more yea-saying for undesirable than desirable behaviors ($Ms = 13.7$ and 2.9 questions, respectively), and significantly more contradictions for people with low compared with higher IQ scores ($Ms = 20.4$ and 13.4 questions, respectively). The authors concluded that cognitive impairments predispose people to response biases and that social desirability then determines the direction this will take. In reverse wordings, then, people may say “yes” to opposite forms of the same question because they are responding to the topic rather than to the particular form of the question, and those who have less ability in receptive language may be particularly vulnerable to this.

In many cases, people are quite willing to answer in the negative (e.g., Booth & Booth, 1994a, 1996; Lowe & de Paiva, 1988; Rapley & Antaki, 1996); therefore, the task is to identify under which circumstances the problem is more likely to arise. It is clearly unacceptable to ignore this difficulty, particularly because it has been found that people may overestimate the comprehension of people with mental retardation (Bartlett & Bunning, 1997; Tully & Cahill, 1984) and that service workers may often fail to adjust their language accordingly (McConkey, Morris, & Purcell, 1999).

Various steps have been taken to counter the problems of yea-saying. Seligman and her colleagues recommended that yes–no questions not be used at all and that either–or questions be used in preference. Many researchers and questionnaire developers now check for consistency by including reverse wordings of all either–or or yes–no questions at other points in the interview (e.g., Shanly & Rose, 1993) or by using screening questions to exclude those participants who are more likely to offer acquiescent responses (e.g., Burnett, 1989; Helsel & Matson, 1988; Jiranek & Kirby, 1990). However, these procedures involve the problems of nonsense and reverse-worded questions described above and may not reflect the complexity of the question structures found in the tests themselves. Tests using absurd questions assume that yea-saying is based on a desire to please or agree with the interviewer. However, if it appears in response to questions that are grammatically or semantically complex or when the person is uncertain of the answer, then such tests would not be a powerful identifier of people who are prone to this.

An alternative when using yes–no questions is to include all participants in the interview, to ensure question wording is clear and simple, and to offer a “don’t know” option. When an answer is given, follow-up, open-ended questions should be used to elicit examples to check the interviewee’s understanding (e.g., Smylie & Ellsworth, 1997; Szivos-Bach, 1993; Zetlin et al., 1985). If examples are not provided by the interviewee, the answer can be further checked using preplanned alternative wordings or phrasings. This allows the detection of acquiescence caused by participants’ focus on only part of the question and, at the same time, allows participation by respondents whose receptive abilities are adequate for yes–no questions, but whose expressive language does not permit them to give examples (Prout & Strohmer, 1994). For interviewees who have greater receptive than expressive abilities, Booth and Booth (1996) suggested strategies by which a narrative might be constructed, such as through the elimination of alternatives and the development of questions by adapting to...
previous responses. However, the validity of this method depends on the interviewer being able to check both that the person does understand the questions as intended and that the person is willing to reject false suggestions. This may prove difficult in practice, and further research needs to be done on this question.

When using reverse wording for all questions (Stancliffe, 1995), the participant may infer that he or she is giving the wrong answer and thus produce more inconsistency (Rapley, 1995). If reverse wording is to be used, it may be helpful to make clear to the interviewee at the outset that this is going to be done and that it does not mean that the original answer was wrong. Alternative phrasings and requests for further detail need to be scripted (e.g., Lindsay & Michie, 1988; Moss et al., 1997) and developed with as much attention as other items. The dangers of spontaneous paraphrasing were illustrated by Antaki (1998), who gave examples of the difficulties that interviewers encountered when they tried to rephrase questions to elicit scoreable responses on a multiple-choice questionnaire. Because of the need to simplify questions, the threshold at which a positive response was scored was often lowered, with the result that overall scores became inflated.

**Multiple-Choice Formats**

Multiple-choice formats used in questionnaires for the general population are usually presented in written form. Because many people with mental retardation cannot read, multiple-choice formats in this population are usually presented orally, although visual aids are also sometimes used. Orally presented multiple-choice formats would present difficulties in the general population but may be even more problematic for many people with mental retardation because of the length of the question, the high memory load required of the respondent (Kahzems, 1985; Reynolds, 1979), and the difficulties in either fitting the response given into one of the categories (Malik et al., 1991; Zetlin et al., 1985) or checking that all the options have been understood (Edmonson, McCombs, & Wish, 1979; Smyley & Ellsworth, 1997). These formats also often involve judgments of degree or frequency. Antaki and Rapley (1996) showed vividly the difficulties of using multiple-choice formats when interviewing people with mental retardation. Using transcripts of quality-of-life interviews, they demonstrated how questions became transformed through the interviewer’s attempts to paraphrase, the need to introduce items, and the need to offer a list of response alternatives. Answers also became transformed through the attempts of the interviewer to fit them into predetermined categories. The examples those researchers gave illustrated that assigning responses to one of the options given in such verbally administered, multiple-choice questionnaires is often a haphazard process.

Difficulties in administering multiple-choice formats were also found in a small study by Malik et al. (1991) in which yes–no, open-ended, and multiple-choice questions were compared. Six people with mild and moderate mental retardation who were recruited from special recreation programs participated in a 287-item interview on social integration and recreation participation. These authors found that test–retest reliabilities were lowest on the multiple-choice questions (58.6%) compared with yes–no and either–or questions (71.8%). Similarly, in an experimental study in which police officers interviewed witnesses with mild and borderline mental retardation, Tully and Cahill (1984) found that although direct, leading questions resulted in 11% of errors, offering alternative choices accounted for one third of all errors. They recommended that first interviews with witnesses with mental retardation should use open-ended questions to avoid this problem, with subsequent interviews used for checking the reliabilities of these accounts and for probing for details.

Although there is some evidence that multiple-choice formats can be used successfully in educational tests with people with mild and borderline mental retardation (Reynolds, 1979), questions that include a list of verbally presented alternatives are not appropriate for many people with mental retardation. When they are used, it is essential to check the answers by asking for examples or by using other, scripted questions (e.g., Clare & Gudjonsson, 1995; Smith & McCarthy, 1996) and by reporting the proportion of responses that are deemed valid (e.g., Benson & Ivins, 1992; Smyley & Ellsworth, 1997).

One solution that is possible in some circumstances is to break the question into two stages. Rybolt (1969) described the use of verbally presented semantic differentials with 73 adolescents classified as having educable mental retardation. This involved presenting the two extremes as an either–or question (e.g., fast or slow), followed by asking whether the item was “a little” or “very fast.” Test–retest reliabilities (up to 3 weeks) for adjective pairs were found to be between .31 and .82. Rybolt also found a significant negative correlation ($r = -.39$) between IQ and levels of response inconsistency.

When response options can be clearly represented in pictorial form and the options are easy to differentiate, pictures can increase responsiveness and understanding (Fox, Faw, Taylor, Davis, & Fulia, 1993; March, 1992; Sigelman & Budd, 1986; Wadsworth & Harper, 1991). Pictures have been found to be less useful when they represent less concrete concepts or when they illustrate quite subtle distinctions (e.g., Sigelman, Schoenrock et al., 1981; Wiggosh & Barry, 1983).

**Open-Ended Questions**

The use of open-ended questions can produce low levels of responsiveness (Sigelman, Budd et al., 1981a; 1981b). Dent (1986) carried out a study with 23 children (aged 8–11) with mild mental retardation who were asked to report on a staged event involving a man who came into their classroom and demonstrated some toys. She found that more detail was produced with specific questions (e.g., “What color was the man’s hair?”), followed by general questions (e.g., “What did the man look like?”), and least detail with free recall ($M_s = 18.8, 9.4, and 4.3$ points of information, respectively). However, the proportion of errors was also highest with specific questions and least with general questions (mean accuracy scores = 64% and 92.7%, respectively)—for similar results, see Perlman et al. (1994). Recently, cognitive interview techniques have been used for witness questioning. These use a range of open-ended questions to enable the interviewee to remember an incident in several different ways (e.g., in reverse order or from another person’s perspective) and have been found to produce greater recall of details (for a review, see Kebbell & Hatton, 1999).

Although open-ended questions might be problematic for some people, many people with mental retardation are able to respond to this type of question, particularly if the questions avoid the types
of phrasings and concepts described above. Open-ended questions have been used successfully in a number of research projects (e.g., Booth & Booth, 1994a, 1994b; Dudley et al., 1997; Edmonson & Wish, 1975; Finlay & Lyons, 2000; Neumayer & Bleasdale, 1996; Shalby & Rose, 1993). Smyley and Ellsworth (1997), in a study of clients’ views of service provision, found that problems of contradictory responses and checking for understanding, which were frequent when using closed questions, were minimized when open-ended questions were used. Rodgers (1999) reported the use of pictures as a prompt in conversational-style interviews to facilitate participation by those who had difficulties in responding to verbal questions.

Quantitative information can also be produced from open-ended questions. Cattermole et al. (1990), in a study of the success of people’s moves to the community from institutions, used a semi-structured interview in which the interviewer had a list of general themes to pursue, such as social life, friends, privacy, and choice. The analysis involved counting the evaluative statements in each category and comparing those made before moving from the hospital to those made after moving. Frequency counts from content analysis were also used by Finlay and Lyons (2000) in a study of social comparisons and by McEvoy (1989), who, investigating people’s understandings of death, used judges to rate people’s answers.

If one is concerned with the point of view of the respondent, in all its complexity, and the researcher does not believe that he or she can second-guess the possible responses, then open-ended questions should be used, despite the difficulties of using such techniques with those who have more severe language difficulties (Bercovici, 1981; Booth & Booth, 1994b; Schurr et al., 1970; Taylor & Bogdan, 1981). Prout and Strohmer (1994) pointed out the importance of considering the expressive–receptive language dichotomy and suggested that a more conversational, open-ended style is better for those with good expressive skills, whereas a more directive style is appropriate for those with more limited expressive abilities (e.g., Booth & Booth, 1996). However, it is important to be aware that the potential for researchers’ imposing their own interpretations and preoccupations is greater when participants are less articulate (Goodley, 1996).

**Answers That Appear Irrrelevant**

If the respondent is preoccupied with certain concerns or problems, he or she may keep returning to these as topics (Biklen & Moseley, 1988; Zetlin et al., 1985). In an attempt to use the Coopersmith Self-Esteem Inventory and the Piers-Harris Self-Concept scale with people with mild to severe mental retardation (mean IQ = 51, SD = 10), Zetlin et al. (1985) found that answers were often incoherent or contradictory or turned into stories or examples that did not seem relevant. Answers that do not seem designed to answer the question might arise for reasons other than a lack of understanding, such as a lack of inhibitory control over previous topics (i.e., when the person finds it difficult to change quickly from one topic to the next—Dempster & Corkill, 1999) or because the interview is proceeding too quickly. In these cases, the question should be repeated later in the interview.

**Psychometric Properties and Target Population**

Many of the statistics that are reported for new or modified self-report questionnaires for people with mental retardation refer to the reliability rather than the validity of the questionnaire in question (e.g., Dagnan et al., 1994; Eysenck, 1965; Flynn et al., 1985; Heal & Chassady-Rusch, 1985; Reiss & Benson, 1985; Reynolds & Miller, 1985; Riggen & Ulrich, 1993; Sinnott-Oswald, Gilner, & Spencer, 1991). Because of the problems described above, researchers should beware of assuming that reliable questionnaires are also valid in this population. Sigelman, Schoenrock, et al. (1981) pointed out that adequate test–retest reliabilities might be due to consistencies in response biases. In addition, many new questionnaires report test statistics from small or vaguely specified samples (e.g., Dagnan et al., 1994; Flynn et al., 1985; Foxx et al., 1993; Heal & Chassady-Rusch, 1985; Nooe, 1977), and researchers using such questionnaires need to investigate their properties further before using them with confidence.

Studies that have reported acceptable reliability statistics for self-report questionnaires have often found that validity statistics do not provide unequivocal support for the instruments (e.g., Chassady-Rusch et al., 1992; Harner & Heal, 1993; Kazdin et al., 1983; Lusky & Benson, 1979; Rojahn et al., 1994). For example, correlations of modified self-report depression scales and the PMRA depression subscale with each other and with informant scales have been inconsistent (e.g., Helsel & Matson, 1988; Kazdin et al., 1983; Senatore et al., 1985). Such inconsistencies might be explained by the criterion scales lacking validity (particularly if these are also self-report), by the use of informants who may have different perspectives (Aman, 1991), or by the invalidity of the scale itself. Examples of the contradictory validity statistics often found in self-report questionnaires in this population are given in Table 1.

A more rigorous approach to questionnaire development would overcome some of the problems of interpretation found in such validity studies. After the content validity is addressed (see earlier in this article), researchers should carry out test refinement procedures (Clark & Watson, 1995; Smith & McCarthy, 1996; see Thorin et al., 1988, for an example in a questionnaire for people with mental retardation) and explicitly address issues of convergent and discriminant validity (Foster & Cone, 1995). Test refinement, which involves collecting large enough samples for psychometric statistics, is likely to be more time-consuming in this population because instruments will usually require individual, oral administration.

Because the validity of many self-report measures has not been demonstrated in this population, convergent and ecological validity should be investigated using third-party accounts (although these also entail problems of validity), behavioral measures, direct measures of the construct, or criterion-related indices. For example, Schalock et al. (1989), in developing the Quality of Life Questionnaire, used validity indices such as staff versions of the questionnaire, goodness-of-fit ratings (ratio of individuals’ behavioral capabilities to the performance requirements of the service setting), measures of staff involvement, and criteria such as type of residential placement. Discriminant validity should be assessed by reporting correlations with measures of response sets, such as social desirability, lying, and yea-saying (Foster & Cone, 1995), as well as IQ scores (Kazdin, 1995).
### Table 1

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reference</th>
<th>Sample</th>
<th>Validity measures</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIMRA (depression scale: yes-no format)</td>
<td>Kazdin et al., 1983</td>
<td>110 adults (96% with borderline to moderate mental retardation)—screening test for acquiescence</td>
<td>Significant correlations with modified BDI, $r = .33$, and Hamilton RSD, informant scale; $r = .2$. No relation to modified ZDS, MMPI depression subscale, and PIMRA depression subscale (informant version). No relation to IQ or level of retardation. Women scored significantly higher than men, $p &lt; .05$.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Matson et al., 1984</td>
<td>As above</td>
<td>Factor structure different for informant and self-report versions of total PIMRA. No affective or depression factor found for self-report version.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Senatore et al., 1985</td>
<td>As above</td>
<td>Those diagnosed on basis of depression subscale had lower adjustment scores ($M = 3.2$ vs. $2.2$) but no differences on other PIMRA subscales, the Social Performance Scale (informant scale), modified BDI, or modified ZDS.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aman et al., 1986</td>
<td>160 adults (borderline to moderate mental retardation) based in community (95) or institution (65)</td>
<td>No relation to sex or level of retardation. Different factors found compared with the results of Matson et al. (1984) for total PIMRA scale. No affective or depression factor found.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Watson, Aman, &amp; Singh, 1986</td>
<td>Same as Aman et al. (1986)</td>
<td>Different factors found for informant version of total PIMRA compared with self-report version. Significant correlations with modified BDI ($r = .38$) and with ZDS ($r = .26$). No relation to receptive vocabulary or level of retardation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Helsel &amp; Matson, 1988</td>
<td>99 adults (93% with borderline to moderate mental retardation); screened for acquiescence</td>
<td>Significant positive relation to neglect as a child ($r = .53$). No significant relation to a range of demographic variables (e.g., IQ, education, marital status, work history).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tymchuk, 1993</td>
<td>31 mothers with mental retardation (mean IQ = 68)</td>
<td>Participants with diagnosed psychopathology scored significantly higher ($M = 2.7$ vs. $0.8$). Significant correlation with informant version ($r$ not reported). No relation to age, sex, or IQ.</td>
<td></td>
</tr>
<tr>
<td>PIMRA (depression scale—Dutch)</td>
<td>van Minnen, Savelsberg, &amp; Hoogduin, 1994</td>
<td>79 adults with mild mental retardation in Holland</td>
<td>Participants with diagnosed psychopathology scored significantly higher ($M = 2.7$ vs. $0.8$). Significant correlation with informant version ($r$ not reported). No relation to age, sex, or IQ.</td>
<td></td>
</tr>
<tr>
<td>Modified Zang Depression Scale (4-point frequency scales using visual-aid bar graph)</td>
<td>Kazdin et al., 1983</td>
<td>110 adults (96% with borderline to moderate mental retardation); screened for use of 4-point scale using questions requiring answer of &quot;never&quot; or other three points (grouped)—see Senatore et al., 1985</td>
<td>Those diagnosed with psychopathology had significantly lower scores ($p &lt; .01$). Significant correlations with modified BDI ($r = .59$), PIMRA total psychopathology score ($r = .25$), and IQ ($r = -.42$). No significant correlation with MMPI depression scale, PIMRA depression subscale (self-report or informant), or Hamilton RSD (informant). High and low scorers differed significantly on BDI ($M = 21.5$ vs. $11.4$), PIMRA total ($M = 21.7$ vs. $17.4$), and Hamilton RSD ($M = 18.9$ vs. $13.9$) but not on PIMRA depression scales (self-report or informant). Those identified as depressed on informant version of PIMRA scored significantly higher ($M = 38.3$ vs. $30.9$).</td>
<td></td>
</tr>
</tbody>
</table>
**Table 1 (continued)**

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reference</th>
<th>Sample</th>
<th>Validity measures</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified Zung Depression Scale</td>
<td>Dagnan &amp; Sandhu, 1999</td>
<td>43 adults with mental retardation; British Picture Vocabulary Scale (mean score = 16.6, range = 6-28)</td>
<td>Significant negative correlation with modified Rosenberg Self-Esteem Scale ($r = .79$) and BDI ($r = .73$). Agreement for clinical diagnosis with BDI = 91%. Scores significantly higher than control sample ($M = 40.38$ vs. 26.0).</td>
<td>Test–retest = .75.</td>
</tr>
<tr>
<td>Unmodified Zung Depression Scale</td>
<td>Prout &amp; Schaefer, 1985</td>
<td>21 adults with mild mental retardation</td>
<td>Significant correlations with unmodified Depression Adjective Checklist ($r = .79$) and BDI ($r = .73$). Agreement for clinical diagnosis with BDI = 91%. Scores significantly higher than control sample ($M = 40.38$ vs. 26.0).</td>
<td></td>
</tr>
<tr>
<td>Modified Beck Depression Inventory (4-point frequency scales using visual-aid bar graph)</td>
<td>Kazdin et al., 1983</td>
<td>110 adults (96% with borderline to moderate mental retardation). Screened for use of 4-point scale using questions requiring answer of “never” or “other three points” (grouped)—see Senatore et al., 1985</td>
<td>Significant correlations with IQ ($r = -.47$), modified ZDS ($r = .59$), modified MMPI depression scale ($r = .25$), PIMRA depression (self-report; $r = .33$) and total scales ($r = .57$), and Hamilton RSD (informant $r = .24$). No significant relation to PIMRA depression or total scales (informant), gender, or level of retardation. High and low scorers differed significantly on ZDS (mean scores = 37.8 vs. 26.1), MMPI depression scale ($M = 26.0$ vs. 22.9), PIMRA self-report depression ($M = 3.2$ vs. 1.8) and total ($M = 22.8$ vs. 12.8) scales, and Hamilton RSD ($M = 18.1$ vs. 12.3) but not on PIMRA informant total or depression scales. Those identified as depressed on informant version of PIMRA scored significantly higher ($M = 22.1$ vs. 15.0).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Helsel &amp; Matson, 1988</td>
<td>99 adults (93% with borderline to moderate mental retardation); screening test not described</td>
<td>Significant correlations with modified ZDS ($r = .35$) and PIMRA depression (self-report; $r = .38$). No significant relation to Hamilton RSD, PIMRA informant depression or total scores, PIMRA self-report total scores, Social Performance Survey (self-report or informant), receptive vocabulary, level of retardation, or gender.</td>
<td></td>
</tr>
</tbody>
</table>
INTERVIEWING PEOPLE WITH MENTAL RETARDATION

Table 1 (continued)

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reference</th>
<th>Sample</th>
<th>Validity measures</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmodified Beck Depression Inventory</td>
<td>Prout &amp; Schaefer, 1985</td>
<td>21 adults with mild mental retardation</td>
<td>Significant correlations with unmodified Depression Adjective Checklist ($r = .74$) and unmodified ZDS ($r = .73$). Agreement for clinical diagnosis with ZDS = 91%. Scores significantly higher than control sample ($M = 19.43$ vs. 4.97).</td>
<td>Alpha = .90. Average item-total correlation = .23. Test-retest correlation = .63 (11 weeks).</td>
</tr>
<tr>
<td>Self-Report Depression Questionnaire—most items 3-point scales</td>
<td>Reynolds &amp; Baker, 1988</td>
<td>89 adults—IQ range, 35–75 (mean = 59.11)—screened for acquiescence</td>
<td>No relation to gender, age, IQ, or score on acquiescence test (after those who failed on screening test removed from sample). Significant correlation with Hamilton clinical interview for depressive symptomatology ($r = .65$).</td>
<td>Little agreement in diagnosis with both Reiss Screen for Maladaptive Behavior depression scale—informant version ($k = 0.06$) and Diagnostic Interview for Children and Adolescents ($k = 0.07$).</td>
</tr>
<tr>
<td></td>
<td>Rojahn et al., 1994</td>
<td>29 adults (mild or moderate mental retardation)—screened for use of 3-point scales</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. BDI = Beck Depression Inventory; Hamilton RSD = Hamilton Rating Scale for Depression; PIMRA = Psychopathology Instrument for Mentally Retarded Adults; ZDS = Zung Depression Scale.

Instruments developed for the general population are unlikely to be appropriate for people with mental retardation (Gowans & Hulbert, 1983; Schurr, Joiner, & Towne, 1970). However, studies frequently use them without demonstrating their validity for this group. This can be found in the assessment of self-concept (e.g., Chiu, 1990; Jiranek & Kirby, 1990; Thompson, Lampron, Johnson, & Eckstein, 1990; Tymchuk, 1991) and psychiatric symptoms (e.g., Manikam, Matson, Coe, & Hillman, 1995; Prout & Schaefer, 1985; Reiss & Benson, 1985). Studies should not report psychometric statistics found for the general population because they are unlikely to be applicable to the groups they are studying (e.g., Calhoun, Whitley, & Ansolabehere, 1978; Chassin, Stager, & Young, 1985; Reynolds & Miller, 1985).

Factor structures found in samples of people with mental retardation may differ from those found in the general population. This is one reason that existing instruments may be inapplicable and creates problems for new or modified questionnaires that are based on factor structures found in the general population (e.g., Riggen & Ulrich, 1993; Szivos-Bach, 1993). For example, Silon and Harter (1985) used the Perceived Competence Scale with a sample of 126 children with mental retardation (IQ range, 55 to 85) in both segregated and mainstream schools and found that instead of the usual four factors, only two factors were present, and there was no global self-worth factor (see Rich, Barckowski, and Witmer, 1979, for a similar example using the Piers-Harris Children's Self-Concept Scale). Matson et al. (1984) also found different factors in the self-report compared with the informant version of the PIMRA (see also Aman, Watson, Singh, Turbott, & Wilsher, 1986). In their study, with people with severe to borderline mental retardation who were users of mental health services, the self-report version yielded two factors (labeled anxiety and social adjustment), whereas the informant version yielded three (affective, somatoform, and psychosis). Because factor structures may be different for informant compared with self-report versions, psychometric statistics should be reported separately for these two groups. Factors or reliability statistics found in mixed samples (e.g., Manikam et al., 1995; McGrew, Bruininks, Thurlow, & Lewis, 1992) should be viewed with caution. Given the problems of validity identified with many self-report questionnaires, however, differences in factor structures that have been found should be considered only as suggestive of differences in underlying constructs.

It is possible that people with mental retardation may be too heterogeneous in terms of personal history and linguistic and cognitive abilities for any single questionnaire to be valid for the whole population. Reliability and validity statistics reported in studies may not therefore be informative for other samples. For example, Prout and Schaefer (1985) used unmodified self-report measures of depression (BDI, ZDS, and the Depression Adjective Checklist) and found high intercorrelations ($r = .73$ to .79). However, there were only 21 participants in this study, all of whom had mild mental retardation. Given the difficulties identified with the questions in such questionnaires, it cannot be assumed that such correlations would be found with people with moderate mental retardation.

Many studies use screening devices or select samples based on vague criteria, making it difficult to judge the applicability of the questionnaire for other samples. Studies often select participants based on a judgment of their communicative skills. For example, Szivos-Bach (1993), in her study of stigma and self-concept in 50 students with mild mental retardation, excluded people who demonstrated that they could not understand the questions asked (for example by replying 'yes' or 'no' to an open-ended question) (p. 220). Aman et al. (1986), in their exploration of the reliability and factor structure of the PIMRA (self-report), included people in the study only when they were 'judged capable of understanding and responding to questions within the PIMRA' (p. 1072). The criteria for this judgment were not described in this paper. Similar vague criteria for selection are found in a number of studies that used self-report questionnaires (e.g., Chadsey-Rusche et al., 1992; Moss...
et al., 1997; Smith & McCarthy, 1996; Stancilfe, 1995; Voelker et al., 1990). The danger is that instruments might have adequate reliability and validity statistics because they are based on a highly selective sample of those with the greatest communicative abilities. When selection criteria are not given or are vague, it is difficult for others to judge whether the obtained scores can be interpreted correctly. Notable exceptions to this are the Emotional Problems Scale (Strohmer et al., 1994), which is for people with mild and borderline mental retardation, and the Eysenck-Withers Personality Inventory (Eysenck, 1965), which was developed for people whose IQ is in the range of 50 to 80.

The degree of verbal ability needed to answer certain types of questions and participate in particular questionnaires needs to be more clearly addressed (Dagnan et al., 1994; Smith & McCarthy, 1996; Smyly & Ellsworth, 1997). A test of verbal competence, such as morphosyntactic ability (Abbeduto & Hagerman, 1997), would be useful for determining the receptive language necessary for certain questionnaires. One example of a questionnaire that assesses auditory comprehension is given by Beaumont, Marjori-banks, Flury, and Lintern (1999). This test includes a range of yes–no questions at different levels of sentence complexity. More specific tests of conceptual understanding, vocabulary, or the ability to use particular response formats (for examples, see Prosser & Bromley, 1998) might be needed for specific instruments.

Finally, some questionnaires can be completed as either self-reports or by informants, such as the PIMRA, the PAS-ADD (Moss et al., 1993, 1997), and the Interview of Community Adjustment (McGrew et al., 1992). It is important in these cases to specify clearly when the self-report version can be used. For example, the PAS-ADD involves a three-tier interview (Moss et al., 1993), each stage involving an assessment of whether the participant can move to the next stage. In a study of elderly people with moderate and severe mental retardation, 40 of 105 people were excluded at the background information stage, and only 14 people completed the interview (Patel et al., 1993). A positive diagnosis is given if either the informant or the self-report version indicates one, but the agreement in diagnosis between the two versions has been found to be rather low. Patel et al. (1993) found that for those who obtained a diagnosis, agreement between the two methods occurred in only one quarter of cases (see also Moss, Prosser, Ibbotson, & Goldberg, 1996; Moss et al., 1997). The authors stated that they have no framework for judging the relative validity of either source, and so validity statistics were based on the combined diagnosis. Despite this problem, the technique might increase the detection rate of psychiatric disorders because it can overcome some of the problems of either respondents with mental retardation who are unable to answer the questions or informants who lack the required information (e.g., about subjective phenomena).

Suggestions for Interviews and Questionnaire Development

A summary of the difficulties in conducting interviews and using self-report questionnaires with people with mental retardation is presented in Table 2. It should be noted that the difficulties described in this paper do not apply to all people with mental retardation, many of whom have few problems with interview questions. For people with borderline and mild mental retardation in particular, many questionnaires developed for the general population may be valid, although this must always be demonstrated rather than assumed. Although it would be wrong to oversimplify questions where it is not necessary, standardized questionnaires developed for widespread use should, however, take note of these difficulties. It should also be acknowledged that certain constructs may be too conceptually complex to be assessed with many people with mental retardation. The suggestions given below, therefore, are aimed at increasing the inclusiveness of interviews and the validity of self-report questionnaires. It is not proposed that all suggestions should be used with all interviewees.

The minimum number of words should be used, and ambiguous or complex phrasings should be avoided. These include structures with modifiers, passive phrasings, and questions that are reversible. Interviewers should always be aware that respondents might be responding to the topic rather than to the particular form of the question. Content such as direct comparisons, judgments of frequency and quantity, and the perceptions of third parties should also be avoided. Comparisons can be split into two parts, and significant events can be used as markers to judge time periods. Rather than asking people to generalize or answer in the abstract, questions may be more successful when situated in specific contexts or events from the person’s own life. It will often be necessary to check how the question was understood by asking for examples or using alternative phrasings (scripted), particularly for yes–no and multiple-choice questions. When probes or rephrasings are to be used, interviewees should be informed at the start of the interview that these are included in the interview as checks and do not mean that their original answer was wrong.

Plenty of time should be allowed for interviews. The administration of self-report questionnaires should not follow the same procedure as with the general population, in which each item elicits a single response and the next item is then addressed. Because of the need for rephrasing and probing, several question-and-answer turns may be necessary for each item to check how the question was understood, to find a phrasing the person understands, and to allow for perseveration and unclear responses. It is better to build such options into the questionnaire rather than require the interviewer to come up with them spontaneously. When probing produces conflicting information, the interviewer should record both responses and discuss them with colleagues after the interview, rather than attempt to make an arbitrary judgment on the spot. In some cases, a consistent, scoreable answer may not be possible.

The difference between expressive and receptive abilities should be recognized because professionals may often overestimate the comprehension of people who appear to have good expressive language abilities. Further research is also needed on techniques of obtaining information from those who can understand questions but have limited expressive abilities. For this group, the problems of checking yes–no responses for understanding are particularly acute.

Tully and Cahill (1984), in their report on police interviewing, suggested an introductory period during which responses to questions for which the answer is already known by the interviewer can be elicited. The interviewer can check how easily the interviewee will agree to suggestions, how he or she exhibits uncertainty, and assure him or her that it is all right to say that he or she does not
Table 2
Summary of Difficulties and Possible Actions

<table>
<thead>
<tr>
<th>Problem area</th>
<th>Specific problems</th>
<th>Possible action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question content</td>
<td>Quantitative judgments</td>
<td>Avoid Likert scales and questions of degree and frequency. Use pretest screening questions about concrete events for which frequency is known.</td>
</tr>
<tr>
<td>Time questions</td>
<td>Use significant events as markers. Ask about each element separately.</td>
<td>Use concrete situations or events. Allow that people may not be able to make generalized judgments.</td>
</tr>
<tr>
<td>Socially reflexive questions</td>
<td>Check meaning of answer.</td>
<td></td>
</tr>
<tr>
<td>Abstract or general concepts</td>
<td>Check meaning of answer.</td>
<td></td>
</tr>
<tr>
<td>Psychiatric symptoms</td>
<td>Avoid symptoms difficult to understand or describe. Elicit item content from appropriate populations. Beware of basing scales on those for general population.</td>
<td></td>
</tr>
<tr>
<td>Irrelevant content</td>
<td>Be aware of the difficulties. Ask about specifics rather than generalities. Stress that information will not be shared with carers.</td>
<td></td>
</tr>
<tr>
<td>Labels</td>
<td>Check understanding.</td>
<td></td>
</tr>
<tr>
<td>Sensitive content</td>
<td>Be aware of the difficulties. Ask about specifics rather than generalities. Stress that information will not be shared with carers.</td>
<td></td>
</tr>
<tr>
<td>Question phrasing</td>
<td>Negative wordings</td>
<td>Avoid adding no and not to positive phrasings. Use negative form of words.</td>
</tr>
<tr>
<td>Modifiers</td>
<td>Avoid modifiers, particularly at end of sentences. Check meaning.</td>
<td></td>
</tr>
<tr>
<td>Subject-object confusion and passive phrasings</td>
<td>Be aware of questions where this is possible—check meaning.</td>
<td></td>
</tr>
<tr>
<td>Question not understood</td>
<td>Keep question structure simple, avoid technical vocabulary. Write alternative phrasings and probes into questionnaires.</td>
<td></td>
</tr>
<tr>
<td>Response format</td>
<td>Yes–no questions</td>
<td>Avoid modifiers and complex question structures. Include “don’t know” option. Check meaning by asking for examples and probing further (use scripted probes). Have preinterview to check how person exhibits uncertainty or responds to false suggestions.</td>
</tr>
<tr>
<td>Multiple-choice format</td>
<td>Break down into two either–or stages. Use pictures only if meaning is clear.</td>
<td></td>
</tr>
<tr>
<td>Understanding or classifying responses</td>
<td>Avoid Likert scales and offering multiple options.</td>
<td></td>
</tr>
<tr>
<td>Psychometric properties</td>
<td>Factor structure</td>
<td>Do not assume that this is the same as for the general population. New questionnaires must clearly report sample selection criteria. Use tests of verbal ability.</td>
</tr>
</tbody>
</table>

They also suggested that two officers be involved so that one can check that the other is not driving the interview and that the importance of accuracy be repeatedly stressed. Introductory periods can also be used to check whether interviewees can make estimates of frequency or degree.

Scale developers need to pay more attention to the validation of scales in terms of both content development and convergence with alternative indicators. The dangers of using questionnaires developed for the general population, or modifications of these, need to be recognized because they may be based on inappropriate content or factor structures. Selection criteria should be given in detail so that clinicians and researchers can decide whether the questionnaire is appropriate for their respondents (Aman, 1991). A questionnaire or interview that excludes a lot of people is at least better than none, but only if this is made clear. Scale developers should identify tests of verbal ability that can be used to specify the receptive language necessary for participation. To date, the issue of how to select who can and cannot take a test has rarely been directly addressed, and more research and discussion is needed on this topic. In certain cases, such as psychiatric diagnoses, it might be worthwhile to use both self-report and informant versions and accept whichever provides a positive identification (Moss et al., 1997; Patel et al., 1993).

Finally, questionnaires should allow for missing or uncodeable answers, rather than requiring the interviewer to fill in every item, and the proportion of responses that are deemed valid should be reported. Yes–no questions should also include “don’t know” options. Tape recording is useful, both to piece together narratives that do not closely follow the questions asked and because participants may be difficult to understand at the time because of unclear speech.
Conclusions

This article has reviewed methodological issues in using interviews and self-report questionnaires with people with mental retardation and outlined a number of question structures and contents that have been found to be problematic. We have argued that because many self-report questionnaires include such questions, more attention needs to be paid to establishing the validity of such measures and to clearly defining the population for which the instrument is designed. A range of strategies to enhance the effectiveness of interviewing and using self-report questionnaires in this population was also reviewed. Although the review necessarily focused on problems with interviewing in this population, its purpose was to allow the voices of people with mental retardation to be heard more clearly. It should be stressed that the difficulties described above do not apply to the population as a whole. The category of people with mental retardation is heterogeneous, and many people can participate in interviews and answer self-report questionnaires without these problems. However, researchers and clinicians must be aware of the sources of potential difficulties to promote the inclusion of as many people as possible and to enhance the validity of interviewing techniques.

References


Coelho, R. J., & Saunders, J. L. (1996). Diagnostic implications of dual...
INTERVIEWING PEOPLE WITH MENTAL RETARDATION


persons with a learning disability detained in police custody. A needs
young adults with mental retardation: A seldom used research method.
*Therapeutic Recreation Journal*, 25, 60–75.
depression: Relationships of self-report to intellectual and adaptive
March, P. (1992). Do photographs help adults with severe mental handi-
caps to make choices. *British Journal of Mental Subnormality*, XXXVII, 122–128.
interviews with adults who have mental retardation. *Mental Retarda-
tion*, 35, 75–82.
eties of the Psychopathology Instrument for Mentally Retarded Adults.
*Applied Research in Mental Retardation*, 5, 81–89.
McConkey, R., Morris, I., & Purcell, M. (1999). Communications between
staff and adults with intellectual disabilities in naturally occurring set-
McEvoy, J. (1989). Investigating the concept of death in adults who are
mentally handicapped. *British Journal of Mental Subnormality*, XXXV, 115–121.
Empirical analysis of multidimensional measures of community adjust-
ment for young adults with mental retardation. *American Journal of
Mental Retardation*, 96, 475–487.
McLean, L. K., Brady, N. C., & McLean, J. E. (1996). Reported communica-
tion abilities of individuals with severe mental retardation. *American
Journal of Mental Retardation*, 100, 580–591.
Moss, S., Ibbotson, B., Prosser, H., Goldberg, D., Patel, P., & Simpson, N.
(1997). Validity of the PAS-ADD for detecting psychiatric symptoms in
adults with learning disability (mental retardation). *Social Psychiatry
Moss, S., Patel, P., Prosser, H., Goldberg, D., Simpson, N., Rowe, S.,
& Lucchino, R. (1993). Psychiatric morbidity in older people with mod-
erate and severe learning disability I: Development and reliability of
the patient interview (PAS-ADD). *British Journal of Psychiatry*, 163, 471–
480.
diagnosis of the Psychiatric Assessment Schedule for Adults with Devel-
and informant accounts of psychiatric symptoms in a sample of patients
Neumayer, R., & Blesadale, M. (1996). Personal lifestyle preferences of
people with an intellectual disability. *Journal of Intellectual and Devel-
opmental Disability*, 21, 91–114.
of African American adolescents with mild mental retardation: Preva-
lence depends on measurement method. *American Journal on Mental
Retardation*, 102, 409–420.
people with moderate and severe learning disability II: The prevalence
Perlman, N. B., Ericson, K. I., Eises, V. B., & Isaacs, B. J. (1994). The
devolutionarily handicapped witness: Competency as a function of
disabilities. In E. Emerson, C. Hatton, J. Bromley, & A. Caine (Eds.),
*Clinical psychology and people with intellectual disabilities*. Chichester,
England: Wiley.
community-based mildly mentally retarded adults. *American Journal of
Mental Deficiency*, 90, 220–222.
Prout, H. T., & Stromher, D. C. (1994). Assessment in counseling and
psychotherapy. In D. C. Strother & H. T. Prout (Eds.), *Counseling
and psychotherapy with persons with mental retardation and borderline
intelligence*. Brandon, VT: Clinical Psychology.
escence” of people with learning disabilities. *Journal of Community and
Reiss, S., & Benson, B. A. (1985). Psychosocial correlates of depression in
mentally retarded adults: I. Minimal social support and stigmatization.
*American Journal of Mental Deficiency*, 89, 331–337.
Reynolds, W. M. (1979). The utility of multiple-choice test formats with
mildly retarded adolescents. *Educational and Psychological Measure-
ment*, 39, 325–332.
persons with mental retardation. *American Journal on Mental Retarda-
tion*, 93, 93–103.
Reynolds, W. M., & Miller, K. L. (1985). Depression and learned help-
lessness in mentally retarded and nonmentally retarded adolescents: An
initial investigation. *Applied Research in Mental Retardation*, 6, 295–
306.
validity of the Piers-Harris children’s self-concept scale for a sample of
intermediate-level EMR students enrolled in elementary school. *Educa-
individuals with mental retardation. *Adapted Physical Quarterly*, 10, 42–51.
Rodgers, J. (1999). Trying to get it right: Undertaking research involving
ment methods for depression in mental retardation. *Journal of Autism
and Developmental Disorders*, 24, 305–313.
speech of mildly retarded adults. In K. T. Kernan, M. J. Begab, & R. B.
Edgerton (Eds.), *Environments and behavior: The adoption of mental-
ly retarded persons*. Baltimore: University Park Press.
on the mentally retarded: A review of empirical studies. *Mental Retar-
dation*, 8, 39–43.
(Eds.), *Child and adolescent psychiatry: Modern approaches*. Oxford,
England: Blackwell.
psychopathology of mentally retarded adults. *American Journal of
Mental Deficiency*, 89, 459–466.
INTERVIEWING PEOPLE WITH MENTAL RETARDATION

335

disabilities currently doing work experience: Their satisfaction with
work and wishes for the future. Mental Handicap Research, 6, 250–262.
Shaw, J. A., & Budd, E. C. (1982). Determinants of acquiescence and
naysaying of mentally retarded persons. American Journal of Mental
Deficiency, 87, 108–110.
mentally retarded persons. Rehabilitation Counselling Bulletin, 29, 173–
181.
Asking questions of retarded persons: A comparison of yes-no and
either-or formats. Applied Research in Mental Retardation, 2, 347–357.
When in doubt say yes: Acquiescence in interviews with mentally
retarded persons. Mental Retardation, 19, 53–58.
Sigelman, C. K., Budd, E. C., Winer, J. L., Schoenrock, C. J., & Martin,
Sigelman, C. K., Schoenrock, C. J., Winer, J. L., Spanhel, C. L., Hromes,
interviewing mentally retarded persons: An empirical study. In R. H.
Bruininks, C. E. Meyers, B. B. Sigford, & K. C. Lakin (Eds.), Deinstitu-
tionalization and community adjustment of mentally retarded people.
Washington, DC: American Association of Mental Deficiency.
Sigelman, C. K., Winer, J. L., & Schoenrock, C. J. (1982). The respon-
siveness of mentally retarded persons to questions. Education and Train-
ing of the Mentally Retarded, 17, 120–124.
motivational orientation, and anxiety in segregated and mainstreamed
educable mentally retarded children. Journal of Educational Psychol-
tree Foundation.
Supported and sheltered employment: Quality of life issues among
workers with disabilities. Education and Training in Mental Retarda-
tion, 388, 397.
Slobin, D. I. (1996). Grammatical transformations and sentence compre-
hension in childhood and adulthood. Journal of Verbal Learning and
interview to investigate the attachment-related experiences of adults
with learning disabilities. British Journal of Learning Disabilities, 24,
154–160.
Smith, S. A. (1993). Confusing the terms "guilty" and "not guilty":
Implications for alleged offenders with mental retardation. Psychologi-
cal Reports, 73, 675–678.
explore client views about a change in service provision. British Journal
Sovner, R., & Hurley, A. D. (1986). Four factors affecting the diagnosis of
psychiatric disorders in mentally retarded persons. Psychiatric Aspects
of Mental Retardation Reviews, 5, 45–49.
comparison of self- and staff-reports. American Journal on Mental
Strohmer, D. C., Prout, H. T., & Gorsky, J. (1994). The development of a
personality inventory for adolescents and adults with mild mental retar-
dation and borderline intelligence. Assessment in Rehabilitation and
Exceptionality, 1, 78–89.
Szivos-Bach, S. E. (1993). Social comparisons, stigma and mainstreaming:
The self-esteem of young adults with a mild mental handicap. Mental
Handicap Research, 6, 217–236.
community adjustment. In R. H. Bruininks, C. E. Meyers, B. B. Sigford,
& K. C. Lakin (Eds.), Deinstitutionalization and community adjustment
of mentally retarded people. Washington, DC: American Association of
Mental Deficiency.
Behavior problems in children with the presenting problem of poor
citizenship rights and responsibilities. Research in Developmental Dis-
abilities, 9, 85–92.
Todd, S., & Shearn, J. (1997). Family secrets and dilemmas of status:
Parental management of the disclosure of “learning disability.” Disabil-
dity and Society, 12, 341–366.
version of the Psychopathology Inventory for Mentally Retarded Adults
adults with mental retardation. Mental Retardation, 28, 305–309.
self-report by adults with moderate mental retardation. Journal of the
Association for Persons with Severe Handicaps, 16, 228–232.
Instrument for Mentally Retarded Adults: Psychometric characteristics,
factor structure, and relationship to subject characteristics. Research in
Wehmeyer, M. L. (1994). Reliability and acquiescence in the measurement of
locus of control with adolescents and adults with mental retardation.
Psychological Reports, 75, 527–537.
people with mental retardation? The National Consumer Survey. Mental
Retardation, 33, 111–119.
TMH students. The Mental Retardation and Learning Disability Bulle-
tin, 11, 79–84.
Lakin (Eds.), Deinstitutionalization and community adjustment of men-
tally retarded people. Washington, DC: American Association of
Mental Deficiency.
ment in retarded adults: A micro-analysis of responses. Applied Re-
search in Mental Retardation, 6, 113–125.

Received May 27, 1999
Revision received January 30, 2001
Accepted March 19, 2001