A review of the reliability and validity of Likert-type scales for people with intellectual disability

S. L. Hartley & W. E. MacLean, Jr.

Department of Psychology, University of Wyoming, Laramie, Wyoming, USA

Abstract

Background Likert-type scales are increasingly being used among people with intellectual disability (ID). These scales offer an efficient method for capturing a wide range of variance in self-reported attitudes and behaviours. This review is an attempt to evaluate the reliability and validity of Likert-type scales in people with ID.
Methods Fifty-one studies reporting response rates, response bias, reliability and validity of Likert-type scales among adolescents and adults with ID were reviewed.
Results Low response rates were found among adolescents and adults with moderate to profound ID, when pictorial representations of response alternatives were not employed, and for Likert-type scales with self-descriptive statements. Response bias was evident, particularly among adolescents and adults with moderate to profound ID. Likert-type scales have better reliability and validity among adolescents and adults with borderline IQ or mild ID. Pretests and clarifying questions increase reliability and validity.
Conclusions Likert-type scales should include pictorial representations of response alternatives, a single set of one or two word response descriptors, clarifying questions, and pretests, and are best used with adolescents and adults with borderline IQ to mild ID.

Keywords intellectual disability, Likert scale, mental retardation, review, self-report measures

Introduction

In the last few decades, there has been a proliferation of self-report measures for people with intellectual disability (ID). Self-report measures exist for a variety of domains, including psychopathology, interpersonal skills, stress, happiness, pain and quality of life (Fogarty *et al.* 1997; Helm 2000; Stancliffe 2000; Kellett *et al.* 2003). This trend represents a departure from the historical reliance on informant reports and observational ratings of people with ID (Stancliffe 2000; Schalock *et al.* 2002).

Self-report measures allow researchers access to the private behaviours, subjective perspectives and mental experiences as acknowledged in literature regarding the general population (e.g. Stone *et al.* 2000; Callahan 2001). For the same reasons, information obtained from self-report measures has become essential in ID research. For example, within quality of life research there is agreement that a key facet of well-being and life satisfaction is only captured through the subjective views of people with ID (e.g. Stancliffe 2000; Schalock *et al.* 2002). Within

Correspondence: Sigan Hartley, Department of Psychology, University of Wyoming, Dept. 3415, 1000 E. University Avenue, Laramie, WY 82071, USA (e-mail: shartley@uwyo.edu).

psychopathology research, assessing mental behaviours such as attitudes and appraisals through self-report measures is thought to be necessary for understanding and identifying psychiatric disorders among people with ID (e.g. Beck *et al.* 1987; Nezu *et al.* 1995). Within stress and coping research, the subjective nature of stress and the necessity of directly asking people with ID about mental coping strategies (i.e. cognitive distraction) has been a focus of recent research (e.g. Bramston *et al.* 1999; Hartley & MacLean 2005). Self-report measures are also important to the field of ID because they allow people with ID to have an active role in research. This follows from the general movement that people with ID need to have a greater say in their lives and should be encouraged to voice their concerns and perspectives (Beart *et al.* 2004).

The use of self-report measures among people with ID is not without challenge. Self-report measures require that people with ID understand questions, form responses independent of the interviewer (e.g. suggestibility) or response format (e.g. order), and communicate responses. Self-report measures also require long- and short-term memory skills to recall past behaviour and attend to multiple response alternatives. Research employing self-report measures with people with ID documents difficulties in response rates, response bias, reliability and validity (Sigelman *et al.* 1981; Heal & Sigelman 1995; Finlay & Lyons 2002).

Within the field of ID, self-report measures have largely consisted of yes/no or either/or questions, and a great deal of research has been devoted to evaluating the validity of these measures (e.g. Sigelman *et al.* 1980, 1981, 1983). More recently, there has been an increase in the use of Likert-type scales (e.g. measures typically consisting of a series of statements in which respondents indicate the degree to which they agree or disagree with each statement) among people with ID (e.g. Mindham & Espie 2003; Hartley & MacLean 2005; Kellett *et al.* 2005). This trend may reflect the documented difficulties that people with ID encounter when answering yes/no and either/or questions (e.g. Sigelman *et al.* 1980, 1983; Heal & Chadsey-Rusch 1985; Sigelman & Budd 1986), or the availability of measures developed for the general population that are appropriate in research with people with ID (e.g. Nezu *et al.* 1995; Glenn *et al.* 2003; Kellett *et al.* 2005).

Response rates, the percentage of participants judged as able to appropriately respond to questions, for self-report measures among people with ID decrease with complexity of the response alternative (Sigelman *et al.* 1981, 1982a,b). *Response rates* are higher for yes/no than either/or questions, and both yes/no and either/or questions have a higher *response rate* than open-ended questions (Sigelman *et al.* 1981, 1982a,b). Likert-type scales require the complex task of distinguishing subtle differences in attitudes or behaviours (e.g. 'Some of the Time' vs. 'Often' or 'Always') and thus may be vulnerable to low *response rates*.

The appropriateness of Likert-type scales among adolescents and adults with ID depends on their reliability and validity. A measure is reliable if it can consistently measure the hypothetical behaviour, quality or trait that it is purported to measure. In order for a measure to be valid, it must adequately capture the hypothetical behaviour, quality or trait that it is purported to measure. Although Likert-type scales are widely used among the general population because they offer an efficient method for capturing a wide range of response variance, little is known about whether people with ID can reliably and validly categorize and distinguish subtle differences (e.g. 'None' to 'A Little', 'Medium' and 'A Lot') in their behaviours and attitudes on Likert-type scales. A critical review of the reliability and validity of Likert-type scales among people with ID has yet to be conducted.

The aim of the present review is to evaluate the reliability and validity of Likert-type scales and identify strategies for increasing the ability of people with ID to accurately respond to these scales. It will be argued that Likert-type scales should be used with adolescents and adults with borderline IQ to mild ID, and include pictorial representations of response alternatives, brief response descriptors, pretests and clarifying questions.

Methodology

Studies were identified using the terms 'mental retardation' or 'intellectual disabilities', and 'self-report' through the PsychInfo or MedLine database. Studies reporting on Likert-type scales among adolescent and adults with ID (aged ≥ 11 years) through: (1) *response rate*; (2) *response bias*, (3) *test–retest reliability*;

(4) *internal consistency* (Cronbach's alpha for total scale scores), (5) *concurrent validity*; and (6) *construct validity* were reviewed. Studies reporting *concurrent* and *construct validity* of Likert-type scales were only included if validity was assessed through comparing Likert-type scales with other self-report measures, historical records or observational data. Studies assessing *concurrent* or *construct validity* through comparing Likert-type scales with informant reports were not included. Informant reports cannot directly access the mental processes, private behaviours or subjective perspectives of people with ID. The degree of convergence between informant reports and self-reported Likert-type scales may not reflect validity (Stancliffe 1995).

Results

Based on these criteria, 51 studies published between 1979 and 2005 were reviewed. Independent samples *t*-tests and one-way analyses of variance (ANOVAs) were conducted to determine differences in *response rates* and *response bias* by number of response alternatives, level of ID and presentation factors. *Test–retest* and *internal consistency reliability* and *concurrent* and *construct validity* were rated as strong, moderate or weak based on pre-established criteria (Table 1) used to evaluate psychological self-report measures (Robinson *et al.* 1991).

Response rate

Twenty studies reported on *response rates* for Likert-type scales among adolescents and adults with ID (Sigelman *et al.* 1982b; Sigelman & Budd 1986; Beck *et al.* 1987; Reynolds & Baker 1988; Benson & Ivins 1992; Chadsey-Rusch *et al.* 1992; Rojahn *et al.* 1994; Dagnan & Ruddick 1995; Nezu *et al.* 1995; Stancliffe 1995; Baker & Bramston 1997; Cummins *et al.* 1997; Lunsky & Benson 2001; Kober & Eggleton 2002; Masi *et al.* 2002; Glenn *et al.* 2003; Powell 2003; Bonham *et al.* 2004; Payne & Jahoda 2004; Hartley & MacLean 2005). *Response rates* ranged from 18.8% to 100% (M = 85.06%, SD = 20.25). At the low end of this range, Likert-type scales have poorer *response rates* than those reported for yes/no (68.1% to 84.8%), either/or (62.8% to 65.7%), and open-ended (50.6% to 75.0%) response options (Sigelman *et al.* 1981, 1982b). However, on average, *response rates* for Likert-type scales are comparable with, and often higher than, those reported for yes/no, either/or and open-ended questions.

Table I Evaluative criteria for reliability and validity

Criterion	Strong	Moderate	Weak
Internal consistency	$\alpha \geq 0.80$	$\alpha = 0.60{-}0.79$	$\alpha < 0.60$
Test–retest reliability	≥ 0.90 (1–3 weeks) ≥ 0.80 (1–2 months) ≥ 0.60 (3–12 months) ≥ 0.50 (1 year)	0.89–0.80 (1–3 weeks) 0.79–0.70 (1–2 months) 0.59–0.50 (3–12 months) 0.49–0.40 (1 year)	<0.80 (1–3 weeks) <0.70 (1–2 months) <0.50 (3–12 months) <0. 40 (1 year)
Concurrent validity	$r \geq 0.70$ with at least two measures; or $r \geq 0.80$ with one measure; or measure discriminated between known groups highly significantly	$r = 0.69{-}0.60$ with at least two measures; or $r \geq 0.70$ with one measure; or measure discriminated between known groups significantly	$r < 0.60$ with at least two measures; or $r < 0.70$ with one measure; or measure did not discriminate between known groups
Construct validity	$r \geq 0.60$ with at least two related measures; or $r \geq 0.70$ with one related measure; or measure was significantly different from unrelated measure	$r = 0.50{-}0.59$ with at least two related measures; or $r = 0.69{-}0.60$ with one related measure	$r \leq 0.50$ with related measures; or measure was not significantly different from an unrelated measure.

Criteria adopted from Robinson *et al.* (1991).

Response rate by number of response alternatives

The optimal number of Likert-type scale response alternatives for the general population has been well researched (e.g. Lundy 1970; Matell & Jacoby 1971; Peter 1979; Cox 1980; Flamer 1983). Little is known about the optimal number of response alternatives to maximize *response rates* among adolescents and adults with ID. In the present review, 14 studies reporting *response rates* used Likert-type scales with three response alternatives, 10 studies employed Likert-type scales with four response alternatives, and 4 studies utilized Likert-type scales with five response alternatives. A one-way ANOVA indicated that there was not a significant difference in *response rate* among Likert-type scales with three (M = 82.19%, SD = 20.48), four (M = 88.17%, SD = 22.16), or five (M = 90.38%, SD = 15.95) response alternatives. This suggests that Likert-type scales with as many as five response alternatives can be utilized with adolescents and adults with ID without significant decreases in *response rates*.

Response rate by level of intellectual disability

Response rates in Likert-type scales differed by intellectual functioning. Within studies, *response rates* were 7% to 33.1% higher for participants with borderline IQ to mild ID than for lower intellectually functioning participants (Sigelman *et al.* 1982a; Bonham *et al.* 2004). Further, an independent samples *t*-test indicated that the *response rate* in studies (Beck *et al.* 1987; Nezu *et al.* 1995; Baker & Bramston 1997; Lunsky & Benson 2001; Kober & Eggleton 2002; Masi *et al.* 2002; Glenn *et al.* 2003; Payne & Jahoda 2004; Hartley & MacLean 2005) assessing adolescents and adults with borderline IQ to mild ID (M = 92.70; SD = 11.37) was significantly higher than the *response rate* in studies (Sigelman *et al.* 1982b; Sigelman & Budd 1986; Reynolds & Baker 1988; Benson & Ivins 1992; Chadsey-Rusch *et al.* 1992; Rojahn *et al.* 1994; Dagnan & Ruddick 1995; Stancliffe 1995; Cummins *et al.* 1997; Powell 2003; Bonham *et al.* 2004) including adolescents and adults with moderate to profound ID (M = 78.34, SD = 23.93), $t = 2.68$, $P = 0.011$.

Response rate by presentation factors

The inclusion of pictorial representations of Likert-type scale response alternatives also appears to be related to an increased *response rate* (Sigelman *et al.* 1982a; Sigelman & Budd 1986; Rojahn *et al.* 1994). For example, Rojahn *et al.* (1994) reported that only 76% of adults with mild to moderate ID ($n = 38$) could respond to a 3-point Likert scale when pictorial representations of response alternatives were not presented. In contrast, 100% of participants could respond to Likert-type scale questions when response alternatives were presented with smiling and frowning faces.

There appears to be a lower *response rate* for Likert-type scales that consist of self-descriptive statements (e.g. 'I have as much energy as ever') as opposed to a single set of one or two word descriptors (e.g. 'a lot'). Beck *et al.* (1987) found that several participants within the mild ID range ($n = 26$) who were able to complete the Depression Self-Rating Scale (Birleson 1981), which uses the brief descriptors 'Never', 'Sometimes', 'Most of the Time', were unable to complete the Beck Depression Inventory (BDI; Beck *et al.* 1961), which uses four self-description response alternatives. Self-descriptive statements require adolescents and adults with ID to decipher among statements with subtle differences that are tailored for specific behaviours or attitudes. Each question presents a new set of response alternatives. In comparison, a standard response alternative set with brief descriptions (i.e. 'None', 'A little', 'Medium' and 'A Lot') only requires adolescents and adults with ID to understand the subtle differences among a limited number of simplified response alternatives that they can then apply to a series of questions.

Response bias

Response bias has been well documented in self-report measures with yes/no and either/or response alternatives among people with ID (Sigelman *et al.* 1981, 1982a, 1983; Loper & Reeve 1983; Sigelman & Budd 1986). The present review suggests that *response bias* also occurs in Likert-type scales. Adolescents and adults with ID have a tendency to choose the most positive response alternative in Likert-type scales (Verri *et al.* 1999; Schalock *et al.* 2000; Hartley & MacLean 2005). As shown in Table 2, the 11 studies reporting *response bias* found that between 0% and 50% (M = 13.2, SD = 17.1) of participants demonstrated a response tendency to choose the most positive response alternative. For instance, Hartley &

Table 2 Response bias in Likert-type scales in studies with adolescents and adults with borderline IQ to mild intellectual disability (ID) and studies including adolescents and adults with moderate to profound ID

| Authors (year) | Likert-type scale | | Response bias |
	Response alternatives	Presentation factors	
Borderline IQ to mild ID			
Emerson (2005) (n = 98)	3, 4		0–23% most positive response to all questions.
Fogarty et al. (1997) (n = 238)	4	1, 2, 3	No evidence of acquiescence.
Hartley & MacLean (2005) (n = 99)	4	1, 2, 3, 4	3.3% most positive response for 100% of questions
Schalock et al. (2000) (n = 237)	3	1	2.5% of participants answered >90% of questions with most positive response
Includes moderate to profound ID			
Bonham et al. (2004) (n = 923)			12% gave most positive response to ≥91% questions. No difference by level of ID
Bramston et al. (1999) (n = 459)	4	1, 2, 3	0.9% rated 82–100% of questions with most positive response option.
Chadsey-Rusch et al. (1992) (n = 51)			50% of severe demonstrated response bias to choose most positive response option
Sigelman & Budd (1986) (n = 109)	3	1	Pictures reduced latter-option bias
Sigelman et al. (1982a) (n = 57)		1	No response bias
Stancliffe (1995) (n = 47)	4	2, 4	2.1% gave most positive response to all questions
Verri et al. (1999) (n = 70 151)	2–5	1	37.1%, 27.2% most positive response to all questions

Response alternatives = Number of Likert-type scale response alternatives. Presentation factors: 1 = pictorial representations of response alternatives; 2 = questions rephrased; 3 = clarifying questions; 4, = pretest.

MacLean (2005) reported that 3% of adults with mild ID in their sample choose the most positive response alternative on a 4-point Likert scale for 100% of the questions. Similarly, Schalock et al. (2000) found that 2.5% of the adults with borderline IQ to mild ID answered more than 90% of the questions with the most positive response alternative. This suggests that *response bias* in Likert-type scales is less frequent than rates of acquiescence, typically varying between 15% and 56%, in yes/no questions (Sigelman et al. 1981, 1982a, 1983; Heal & Chadsey-Rusch 1985; Heal & Sigelman 1995) and comparable with *response rates* for choosing the latter alternative, varying between 9.4% and 28.6%, in either/or questions (Sigelman et al. 1981, 1983).

Response bias by level of intellectual disability

The tendency to choose the most positive response alternative, however, is more problematic among adolescents and adults functioning at lower levels of ID. In their sample of 923 adults with ID, of whom 45% had moderate to profound ID, Bonham et al.

(2004) reported that 12% choose the most positive response alternative on a 3-point Likert-type scale for more than 90% of the questions. Verri et al. (1999) excluded 37% of their Italian sample and 27% of their Australian sample of adults with mild to moderate ID because they responded with the most positive alternative for 100% of questions using a 3- to 5-point Likert-type scale. Thus, a tendency to choose the most positive response alternative in Likert-type scales appears to be related to intellectual functioning and is more problematic among adults with lower intellectual functioning.

Response bias by presentation factors

Allowing interviewers to paraphrase and/or expand upon items appears to help adolescents and adults with ID reliably respond to Likert-type scales (Table 2). In the present review, interviewers were instructed to paraphrase and/or expand upon items that were not readily understood by participants in 30% of the studies (6/20) reporting *response rates* (Chadsey-Rusch et al. 1992; Stancliffe 1995; Baker &

S. Hartley & W. MacLean • **Intellectual disability and Likert-type scales**

Bramston 1997; Kober & Eggleton 2002; Masi *et al.* 2002; Hartley & MacLean 2005). The five of these studies reporting on adolescents and adults with borderline IQ to moderate ID (Stancliffe 1995; Baker & Bramston 1997; Kober & Eggleton 2002; Masi *et al.* 2002; Hartley & MacLean 2005), reported high *response rates* (M = 94.5%, SD = 3.6) and/or low *response bias* (M = 1%, SD = 0.4).

Reliability and validity

Overall, the present review suggests that Likert-type scales have adequate reliability and validity among adolescents and adults with ID. As depicted in Table 3, the majority of studies found moderate to strong *internal consistency* (17/20), *test–retest reliability* (7/11) and *concurrent validity* (10/17). Further, almost half (9/20) of the studies found moderate to strong *convergent validity*. Level of ID and the use of pretests and/or clarifying questions appear to increase the reliability and validity of Likert-type scales among adolescents and adults with ID.

Reliability and validity by level of intellectual disability

Almost two-thirds (10/16) of the studies reporting on reliability and validity for Likert-type scales used with adolescents and adults with moderate to severe ID received a rating of weak (Kazdin *et al.* 1983; Senatore *et al.* 1985; Helsel & Matson 1988; Benson & Ivins 1992; Lindsay *et al.* 1994; Rojahn *et al.* 1994; Gullone *et al.* 1995; Cummins *et al.* 1997; Bramston & Fogarty 2000; Powell 2003). In contrast, only a little over one-third (9/23) of studies reporting on reliability and validity for Likert-type scales used with adolescents and adults with borderline IQ to mild ID received a rating of weak (Reiss & Benson 1984; Reynolds & Miller 1985; Luftig 1988; Bramston *et al.* 1993; Nezu *et al.* 1995; Dagnan & Sandhu 1999; Kober & Eggleton 2002; Masi *et al.* 2002; Payne & Jahoda 2004). This suggests that adolescents and adults with borderline IQ to mild ID respond more consistently and accurately to Likert-type scales than adolescents and adults with moderate to severe ID. For instance, Kazdin *et al.* (1983) found a low correlation (0.10) between the depression scale of the Psychopathology Instrument for Mentally Retarded Adults (PIMRA-D) and the BDI-R, a 4-point Likert-type scale of depression. Forty-five per cent of their

sample (*n* = 110) were adults with moderate to severe ID. In contrast, Nezu *et al.* (1995), using a similar sample size (*n* = 107), found a moderate correlation (0.53) between the PIMRA-D and BDI-R among adults with mild ID.

Reliability and validity by presentation factors

The inclusion of pretests to identify and eliminate participants who demonstrate inappropriate or contradictory response tendencies appears important for achieving moderate to strong reliability and validity among adolescents and adults with moderate to severe ID. Table 4 displays the types of pretests that have been used for Likert-type scales among adolescents and adults with ID and the advantages and disadvantages of each. As shown in Table 3, all of the studies incorporating pretests among participants with moderate to severe ID reported at least moderate *internal consistency* or *test–retest reliability* for Likert-type scales (Chadsey-Rusch *et al.* 1992; Cummins *et al.* 1997; Powell 2003). For instance, Powell (2003) included a six-question Likert-type scale pretest, in which three questions required a response of 'never' (e.g. 'You like to eat rotten food') and three required a response of 'some of the time' or 'a lot of the time' (e.g. 'You like money') to identify participants unable to reliably respond to a Likert-type scale. After eliminating participants who demonstrated inappropriate responses in the pretest, Powell found strong *internal consistency* (α = 0.86) for the BDI among adults with mild to severe ID. Pretests thus appear to be an effective strategy for ensuring that adolescents and adults with ID are able to reliably respond to Likert-type scales.

Pretests can also be used to decrease *response bias* through training participants to distinguish among response alternatives. In pretests, participants become familiar and comfortable with response formats while they are provided with feedback on how to use them. For instance, Bromley *et al.* (1998) used a pretest to familiarize and train 20 adults with ID to use response formats denoting location and intensity of pain. In their study, participants were asked to arrange blocks of varying sizes in ascending order and to place the blocks at their appropriate points along a pain ruler, which was later used during actual testing. Further, to prepare participants for identifying pain locations, participants were asked to point out

Table 3 Reliability and validity of Likert-type scales in studies with adolescents and adults with borderline IQ to mild intellectual disability (ID) and studies including adolescents and adults with moderate to severe ID

Authors (year)	Likert-type scale		Reliability		Validity	
	Response alternatives	Presentation factors	Internal consistency	Test–retest	Concurrent validity	Convergent validity
Borderline IQ to mild ID						
Ailey (2000) (n = 27)	3		Moderate	Strong		
Baker & Bramston (1997) (n = 105)	4, 5	1, 2, 3			Moderate	Strong
Beck et al. (1987) (n = 26)	3, 4	3	Strong			
Bramston & Bostock (1994) (n = 221)	4	1, 2	Strong to weak			
Bramston et al. (1993) (n = 28)	4	3, 4	Moderate			
Bramston & Miochie (2001) (n = 31)	5	1				
Dagnan & Sandhu (1999) (n = 43)	3, 4	1	Moderate	Moderate to weak		Weak
Emerson (2005) (n = 98)	4		Strong to moderate		Strong to moderate	
Fogarty et al. (1997) (n = 238)	4	1, 2, 3	Strong		Strong to moderate	Strong
Glenn et al. (2003) (n = 46)	5	1, 4	Strong		Moderate	
Kellett et al. (2005) (n = 225)	5	1		Moderate		
Kellett et al. (2003) (n = 200)	3	1				Moderate
Kober & Eggleton (2002) (n = 172)	5	2	Strong	Weak		Moderate to weak
Luftig (1988) (n = 20)	3, 4	4				
Lunsky & Benson (1997) (n = 50)	5, 5		Strong		Moderate	Weak
Masi et al. (2002) (n = 50)	4, 5	2				Moderate
Nezu et al. (1995) (n = 107)	4, 3, 5	1, 4			Weak	
Novaco & Taylor (2004) (n = 129)	3, 5	2	Strong	Moderate	Strong	Strong
Payne & Jahoda (2004) (n = 38)	4		Moderate	Strong		Weak
Prout & Schaeffer (1985) (n = 21)	3, 5				Strong	
Reiss & Benson (1984) (n = 32)	3		Moderate	Moderate to weak		
Reynolds (1979) (n = 100)	4, 3		Strong			
Reynolds & Miller (1985) (n = 26)						Weak

Table 3 *Continued*

Authors (year)	Likert-type scale		Reliability		Validity	
	Response alternatives	Presentation factors	Internal consistency	Test–retest	Concurrent validity	Convergent validity
Includes moderate to severe ID						
Benavidez & Matson (1993) (n = 25)	3, 4, 5	1			Strong	
Benson & Ivins (1992) (n = 130)	3, 4	1				Weak
Bramston & Fogarty (2000) (n = 147)	4, 3	1, 3	Strong to moderate		Weak	Weak
Chadsey-Rusch et al. (1992) (n = 51)	3	1, 2, 4		Strong to moderate		
Cummins et al. (1997) (n = 59)	2–5	1, 4	Moderate to weak	Moderate to weak		
Gullone et al. (1995) (n = 51)	3, 4	1, 2			Weak	Weak
Helsel & Matson (1988) (n = 99)	3, 4	1, 4			Weak	Weak
Kazdin et al. (1983) (n = 110)	4, 5				Weak	Strong
Lindsay et al. (1994) (n = 67)	4	1				Moderate
Lunsky (2003) (n = 99)	3, 4					
Mindham & Espie (2003) (n = 35)	3, 4	1, 2, 3	Strong	Strong	Strong	
Powell (2003) (n = 120)	4, 5	1, 2, 4	Strong to weak		Weak	Moderate
Reynolds & Baker (1988) (n = 103)	3, 5	1	Strong	Strong	Moderate	
Rojahn et al. (1994) (n = 38)	3, 5				Weak	Moderate
Schloss (1982) (n = 18)	4					
Senatore et al. (1985) (n = 110)	4	4				Weak

Response alternatives = number of Likert-type scale alternatives. Presentation factors: 1 = pictorial representation of response alternatives; 2 = questions rephrased; 3 = clarifying questions; 4 = pretest. Reliability and validity ratings with more than one category reflect studies reporting on more than one Likert-type scale or studies that assessed reliability and validity more than one time.

Table 4 Pretests for Likert-type scales for adolescents and adults with intellectual disability (ID)

Authors (year)	Description of Pretest	Advantage	Disadvantage
Baker & Bramston (1997)	Minnesota Multiphasic Personality Inventory Lie Scale (score ≥ 8)	1) Norms for general population available	1) Does not train for Likert-type scales 2) Does not assess response bias for Likert-type scales
Bromley et al. (1998)	Distinguish between a big and small block, arrange blocks of varying sizes in ascending order, and place blocks at their appropriate points along a ruler (later used in actual testing). Participants were asked to point out various parts of their body and indicate where these same body parts were on a bodymap (later used in actual testing). Participants were corrected up to four times for any item	1) Allows repeated trials to aid learning 2) Parallels Likert-type scale in testing	1) Requires a long time
Chadsey-Rusch et al. (1992)	Questions using same 3-point response scale as in actual test	1) Brief 2) Same Likert scale as used in actual testing	1) Three items may not be sufficient to identify response bias
Cummins et al. (1997); Hartley & MacLean (2005); Verri et al. (1999)	Designate size–order relationships among a set of blocks, relate block size to written scale of size, and place something of known desirability on a written scale of preference. Tasks progressed from a 2- to a 5-point Likert-type scale depending on each participant's ability to master each response format. Pretest was then used to tailor the response format in actual testing	1) Tailoring response options may increase response rate 2) Manipulation of 3-D objects may increase learning	1) Varying number of response options may make analyses difficult 2) Questions may not be comparable to actual testing questions
Glenn et al. (2003)	4 closed-ended simple questions (e.g. 'Have you ever been to the moon?') that required a specific yes or no answer. If answered any incorrectly, excluded	1) Brief	1) Does not train for Likert-type scales 2) Does not assess response bias for Likert-type questions
Helsel & Matson (1988)	3- and 4-point Likert screening ('Is it daytime outside?'). Must answer more than 50% of the screening items. Use of bar graph was demonstrated with modeling	1) Models use of pictures to determine answers	1) Pretest questions irrelevant to questions in actual testing

Table 4 *Continued*

Authors (year)	Description of Pretest	Advantage	Disadvantage
Luftig (1988)	5 practice items using a 5-point Likert-type scale in study	1) Related to questions in actual testing	1) No established criteria for identifying response bias
Lunsky & Benson (2001); Lunsky *et al.* (2002)	Must respond correctly to sample items	1) Brief 2) Related to questions in actual testing	1) Single item may not be sufficient to identify response bias
Mindham & Espie (2003)	Examiner demonstrated how to respond using cue cards and checked understanding by using everyday examples (e.g. 'Do you like to go to the cinema?')	1) Examiner demonstrates use of scales	1) No criteria for identifying response bias
Nezu *et al.* (1995)	5 questions using a 4-point scale ('How often do you see a dog colored green?'). If two or more answered incorrectly, excluded from the study	1) Brief	1) Response bias to improbable questions may not reflect response bias to actual questions
Reynolds & Baker (1988)	Social and Prevocational Information Battery-Form T pretest. Must accurately answer 15 of the 20 yes/no format questions	1) Large number of pretest items may increase ability to identify response bias	1) Does not train for Likert-type scales 2) Does not assess response bias for Likert-type scales
Powell (2003); Senatore *et al.* (1985)	Three questions requiring a response of 'never' (e.g. 'You like to eat rotten food') and 3 requiring a response of 'some of the time' or 'a lot of the time' (e.g. 'You like money'). Required to answer 5 of the 6 questions correctly	1) Accounts for true variation in responses (i.e. 'some of the time' or 'a lot of the time')	1) Response bias to improbable questions may not reflect response bias to actual questions
Stancliffe (1995)	One question ('Do you choose who lives next door?') requiring response of 'no'	1) Brief 2) Questions were relevant to actual testing questions	1) Does not train for Likert-type scales 2) Does not assess response bias for Likert-type scales

various parts of their body and to indicate where these same body parts were on a bodymap, also later used in actual testing (Table 4). If participants responded incorrectly during the pretest, they were corrected up to four times for any item. Thus, this pretest procedure not only was used to identify and eliminate adults with ID who demonstrated *response bias*, but taught adults with ID to use Likert-type scales so that they could be included in actual testing.

Pretests have also been used to reduce *response bias* through determining the complexity of Likert-type scale for which adolescents and adults with ID can reliably respond, and then modifying Likert-type scales in actual testing accordingly (Table 4). Verri *et al.* (1999) and Cummins *et al.* (1997) had adults with ID designate size–order relationships among a set of blocks, relate block size to a written scale of size, and place something of known desirability on a written scale of preference. These tasks first pertained to a two alternative scale ('big' or 'small') and progressed to 3- or 5-point Likert-type scale, depending on each participant's ability to master each response format. This pretest was then used to tailor the Likert-type scale of interest, such that the response format corresponded to the scale complexity mastered by each participant in the pretest. This strategy maximizes response rates by allowing adolescents and adults with ID with varying abilities to respond to Likert-type scales to be included in actual testing and may increase *response rates* among adolescents and adults functioning in the lower range of ID.

Another strategy for increasing validity of Likert-type scales appears to be the inclusion of clarifying questions (Table 3). For instance, Bramston *et al.* (1993, 1999) used the prompt 'Tell me more about it' to ensure that questions were understood and responses appropriately reflected participants' perspectives. All five studies reporting Cronbach's alphas of Likert-type scales that incorporated clarifying questions found moderate to high *internal consistency* (Bramston *et al.* 1993; Baker & Bramston 1997; Fogarty *et al.* 1997; Bramston & Fogarty 2000; Mindham & Espie 2003). Two of these studies included participants with moderate to severe ID, suggesting that clarifying questions may help ensure accurate responding by adolescents and adults in the lower ranges of ID. Studies including clarifying questions reported low rates of *response bias* in the form of choosing the most positive response alternative

(Stancliffe 1995; Fogarty *et al.* 1997; Bramston *et al.* 1999; Hartley & MacLean 2005). Incorporating standardized clarifying questions in the administration of Likert-type scales also may provide a checking system to help ensure responses appropriately reflect true opinions.

Discussion

Likert-type scales offer an efficient method for capturing a wide range of response variance in the self-reported attitudes and behaviours of people with ID. The use of these scales has implications for understanding the private behaviours, mental experiences, and subjective perspectives and attitudes of people with ID. This information is essential for many areas of research in ID. This review is an attempt to evaluate the reliability and validity of Likert-type scales in adolescents and adults with ID.

The present review of 51 studies suggests that *response rates* for Likert-type scales, on average, are comparable with those reported for yes/no, either/or and open-ended questions. It is important to note, however, that the average *response rate* in the present review is not representative of all adolescents and adults with ID. Nearly half of the studies (9/20) reporting *response rates* consisted of adolescents and adults with borderline IQ or mild ID (Beck *et al.* 1987; Nezu *et al.* 1995; Baker & Bramston 1997; Lunsky & Benson 2001; Kober & Eggleton 2002; Masi *et al.* 2002; Glenn *et al.* 2003; Payne & Jahoda 2004; Hartley & MacLean 2005), and only one study included adults with profound ID (Bonham *et al.* 2004). *Response rates* may be markedly lower if adults with moderate to profound ID were represented to a larger degree. Further, many of the studies (8/20) reporting *response rates* required that participants have adequate communication skills and/or be judged by researchers or caregivers as able to understand questions and/or use a Likert-type response format (Rojahn *et al.* 1994; Stancliffe 1995; Baker & Bramston 1997; Lunsky & Benson 2001; Masi *et al.* 2002; Glenn *et al.* 2003; Payne & Jahoda 2004; Hartley & MacLean 2005). The number of potential participants excluded prior to actual testing because of these inclusion criteria was not reported in the majority of studies. *Response rates* for Likert-type scales may be substantively lower when no inclusion criteria are used.

Likert-type scales are prone to low *response rates* among adolescents and adults with moderate to profound ID and when pictorial representations of response alternatives are not employed (Sigelman *et al.* 1982a; Chadsey-Rusch *et al.* 1992; Bonham *et al.* 2004). Pictorial representations of response alternatives may guide adolescents and adults with ID in distinguishing the subtle differences among responses. *Response rates* are also low in Likert-type scales with self-descriptive statements as contrasted with a single set of one or two word descriptors used with a series of questions. Adolescents and adults with ID, and particularly those with moderate to profound ID, demonstrate a tendency to choose the most positive response alternative in Likert-type scales (e.g. Verri *et al.* 1999; Bonham *et al.* 2004).

The present study suggests that as many as five response alternatives can be used in Likert-type scales without decreases in *response rates*. This does not mean that 5-point Likert-type scales are most advantageous for people with ID. Discussions of Likert-type scales more generally, suggest that there are multiple considerations in determining the optimal Likert-type scale (see Hodge & Gillespie 2003). Increasing the number of response alternatives increases the amount of information collected, and subsequently reliability. However, response alternatives should correspond to respondents' actual experiences, and fewer response alternatives may more closely parallel these experiences (Chang 1994). Conclusions regarding the optimal number of Likert-type scale response alternatives cannot be determined until these considerations are examined in people with ID.

Allowing interviewers to paraphrase and/or expand upon question items or response alternatives is related to an increased *response rate* and decreased *response bias* among adolescents and adults with ID (e.g. Fogarty *et al.* 1997; Kober & Eggleton 2002). However, a lack of consistency in presenting questions could introduce variability in responses that could affect study findings. Antaki (1999) warned that this strategy inadvertently biases responses of people with ID in the direction of the expectations of the interviewer. A balance between allowing flexibility in asking questions to maximize *response rates* and maintaining standardized administrations of Likert-type scales is needed. One possible solution may be incorporating scripted paraphrasing of questions and/

or expanded definitions of terms in Likert-type scale questions to adolescents and adults with ID. For instance, Lunsky *et al.* (2002) instructed interviewers to read items word-for-word when administering the Healthy Behaviours Screen, a measure of health-related behaviours and common health complaints, to adults with ID. However, if participants did not understand an item or appeared confused, a scripted operational definition of the item was provided. For example, if participants did not understand 'nausea', the explanation 'when you feel like you are going to throw up or vomit' was read. This strategy maximizes *response rates* by providing expanded definitions of items to people with ID who are confused, yet limits the opportunity for interviewer bias.

Likert-type scales have better reliability and validity among adolescents and adults with borderline IQ to mild ID than among those with moderate to severe ID. Pretests are an effective strategy for increasing reliability and validity, particularly among adolescents and adults with moderate to severe ID (Chadsey-Rusch *et al.* 1992; Cummins *et al.* 1997; Powell 2003). Pretests can be used to identify participants who demonstrate inappropriate or contradictory response tendencies and to increase *response rates* through training adolescents and adults with ID to use Likert-type scales (e.g. Bromley *et al.* 1998; Hartley & MacLean 2005). Pretests can also be used to determine the complexity of Likert-type scales for which each participant can reliably respond and then modify the Likert-type scale in actual testing accordingly (Cummins *et al.* 1997; Verri *et al.* 1999). This strategy may help increase *response rates* among adolescents and adults with moderate to severe ID. Incorporating clarifying questions into the administration of Likert-type scales also appears to be an effective strategy for ensuring that responses reflect the true opinions of adolescents and adults with ID.

One strategy for evaluating the reliability and validity of Likert-type scales among adolescents and adults with ID was used in this review. In order to draw conclusions across studies, criteria for weak, moderate, and strong reliability and validity were established. This critique system does not account for differences in the expected strength of relationships between Likert-type scales and measures of behaviours theoretically related to these scales (e.g. a measure of low social support may theoretically be less strongly related to a Likert-type scale of depression

than a measure of automatic negative thoughts). Low convergent validities may be expected for behaviours thought to only be weakly related to behaviours assessed in the Likert-type scale. In this case, low *convergent validity* is not a sign of poor validity. It is also important to note that we only reviewed published research using Likert-type scales. Much of the research demonstrating unreliable and invalid Likert-type scales may not have been published. Future research will need to continue to investigate the conditions under which Likert-type scales are most appropriate for people with ID. Finally, studies varied in their criteria for evaluating level of ID, and thus categories of borderline IQ, mild, moderate and severe ID may not be consistent across studies.

Within the general population, the merit of Likert-type scales is often questioned (see Hodge & Gillespie 2003). Likert-type scales constitute ordinal-level data, yet responses are typically summed and treated as interval or ratio-level data. The same cautions that are advised in interpreting results from Likert-type scales more generally (e.g. Russell & Bobko 1992; Nanna & Sawilowsky 1998), apply to the field of ID. Future research will also need to investigate potential differences in the reliability and validity of Likert-type scales among people with ID for questions regarding overt behaviours (e.g. crying) vs. abstract feelings or attitudes (e.g. feelings of hopeless).

References

Ailey S. H. (2000) Screening adolescents with mental retardation for depression. *Journal of School Nursing* **16,** 6–11.

Antaki C. (1999) Interviewing persons with a learning disability: how setting lower standards may inflate well-being scores. *Qualitative Health Research* **9,** 437–54.

Baker W. & Bramston P. (1997) Attributional and emotional determinants of aggression in people with mild intellectual disabilities. *Journal of Intellectual and Developmental Disability* **22,** 169–86.

Beart S., Hardy G. & Buchan L. (2004) Changing selves: a grounded theory account of belonging to a self-advocacy group for people with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities* **17,** 91–101.

Beck A. T., Ward C. H., Mendalism M., Mock J. & Erbaugh J. (1961) An inventory for measuring depression. *Archives of General Psychiatry* **4,** 561–71.

Beck D. C., Carlson G. A., Russell A. T. & Brownfield F. E. (1987) Use of depression rating instruments in developmentally and educationally delayed adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry* **26,** 97–100.

Benavidez D. A. & Matson J. L. (1993) Assessment of depression in mentally retarded adolescents. *Research in Developmental Disabilities* **14,** 179–88.

Benson B. A. & Ivins J. (1992) Anger, depression and self-concept in adults with mental retardation. *Journal of Intellectual Disability Research* **36,** 169–75.

Birleson P. (1981) The validity of depressive disorders and the development of a self-rating scale: a research report. *Journal of Child Psychology and Psychiatry* **22,** 73–88.

Bonham G. S., Basehart J., Schalock R. L., Marchand C. B., Kirchner N. & Rumerap J. M. (2004) Consumer-based quality of life assessment: the Maryland Ask Me! Project. *Mental Retardation* **42,** 338–57.

Bramston P. & Bostock J. (1994) Measuring stress in people with intellectual disabilities: the development of a new scale. *Australian and New Zealand Journal of Developmental Disabilities* **19,** 149–57.

Bramston P. & Fogarty G. J. (2000) The assessment of emotional distress experienced by people with an intellectual disability: a study of different methodologies. *Research in Developmental Disabilities* **21,** 487–500.

Bramston P. & Miochie C. (2001) Disability and stress: a study in perspectives. *Journal of Intellectual and Developmental Disability* **26,** 233–42.

Bramston P., Bostock J. & Tehon G. (1993) The measurement of stress in people with an intellectual disability: a pilot study. *International Journal of Disability Development and Education* **40,** 95–104.

Bramston P., Fogarty G. & Cummins R. A. (1999) The nature of stressors reported by people with an intellectual disability. *Journal of Applied Research in Intellectual Disabilities* **12,** 1–10.

Bromley J., Emerson E. & Caine A. (1998) The development of a self-report measure to assess the location and intensity of pain in people with intellectual. *Journal of Intellectual Disability Research* **42,** 72–81.

Callahan R. J. (2001) Thought field therapy: response to our critics and a scrutiny of some old ideas of social sciences. *Journal of Clinical Psychology* **57,** 1251–60.

Chadsey-Rusch J., DeStefano L., O'Reilly M., Gonzalez P. & Collet-Klingenberg L. (1992) Assessing the loneliness of workers with mental retardation. *Mental Retardation* **30,** 85–92.

Chang L. (1994) A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement* **18,** 205–15.

Cox E. P. (1980) The optimal number of response alternatives for a scale: a review. *Journal of Marketing Research* **17,** 407–22.

Cummins R. A., McCabe M. P., Romeo Y., Reid S. & Waters L. (1997) An initial evaluation of the comprehensive quality of life scale–intellectual disability. *International Journal of Disability, Development, and Education* **44,** 7–19.

Dagnan D. & Ruddick L. (1995) The use of analogue scales and personal questionnaires for interviewing people with learning disabilities. *Clinical Psychology Forum* **13,** 21–4.

Dagnan D. & Sandhu S. (1999) Social comparison, self-esteem, and depression in people with intellectual disability. *Journal of Intellectual Disability Research* **43,** 372–7.

Emerson E. (2005) Use of the Strengths and Difficulties Questionnaire to assess the mental health needs of children and adolescents with intellectual disabilities. *Journal of Intellectual and Developmental Disabilities* **30,** 14–23.

Finlay W. M. & Lyons E. (2002) Acquiescence in interviews with people who have mental retardation. *Mental Retardation* **40,** 14–29.

Flamer S. (1983) Assessment of the multitrait–multimethod matrix validity of Likert scales via confirmatory factor analysis. *Multivariate Behavioral Research* **18,** 275–308.

Fogarty G. J., Bramston P. & Cummins R. A. (1997) Validation of the Lifestress Inventory for people with a mild intellectual disability. *Research in Developmental Disabilities* **18,** 435–56.

Glenn E., Bihm E. M. & Lammers W. J. (2003) Depression, anxiety, and relevant cognitions in persons with mental retardation. *Journal of Autism and Developmental Disorders* **33,** 69–76.

Gullone E., Cummins R. A. & King N. J. (1995) Adaptive behaviour in children and adolescents with and without an intellectual disability: relationships with fear and anxiety. *Behaviour Change* **12,** 227–37.

Hartley S. L. & MacLean W. E. (2005) Perceptions of stress and coping strategies among adults with mental retardation. *American Journal on Mental Retardation* **110,** 285–90.

Heal L. W. & Chadsey-Rusch J. (1985) The Lifestyle Satisfaction Scale (LSS): assessing individual's satisfaction with residence, community setting and associated services. *Applied Research in Mental Retardation* **6,** 475–90.

Heal L. W. & Sigelman C. K. (1995) Response biases in interviews of individuals with limited mental ability. *Journal of Intellectual Disability Research* **39,** 331–40.

Helm D. T. (2000) The measurement of happiness. *American Journal on Mental Retardation* **105,** 326–35.

Helsel W. J. & Matson J. L. (1988) The relationship of depression to social skills and intellectual functioning in mentally retarded adults. *Journal of Mental Deficiency Research* **32,** 411–18.

Hodge D. R. & Gillespie D. (2003) Phrase completions: an alternative to Likert scales. *Social Work Research* **27,** 45–56.

Kazdin A. E., Matson J. L. & Senatore V. (1983) Assessment of depression in mentally retarded adults. *American Journal of Psychiatry* **140,** 1040–3.

Kellett S., Beail N., Newman D. W. & Frankish P. (2003) Utility of the Brief Symptom Inventory in the assessment of psychological distress. *Journal of Applied Research in Intellectual Disabilities* **16,** 127–34.

Kellett S., Beail N. & Newman D. W. (2005) Measuring interpersonal problems in people with mental retardation. *American Journal on Mental Retardation* **110,** 136–44.

Kober R. & Eggleton I. R. (2002) Factor stability of the Schalock and Keith (1993) Quality of Life Questionnaire. *Mental Retardation* **40,** 157–65.

Lindsay W. R., Michie A. M., Baty F. J. & Smith A. H. (1994) The consistency of reports about feelings and emotions from people with intellectual disabilities. *Journal of Intellectual Disability Research* **38,** 61–6.

Loper A. B. & Reeve R. E. (1983) Response bias on a locus of control measure by learning-disabled children. *Journal of Abnormal Child Psychology* **11,** 537–48.

Luftig R. L. (1988) Assessment of the perceived school loneliness and isolation of mentally retarded and nonretarded students. *American Journal on Mental Retardation* **92,** 472–5.

Lundy G. H. (1970) Using analysis of variance with a dichotomous dependent variable. An empirical study. *Journal of Educational Measurement* **7,** 263–9.

Lunsky Y. (2003) Depressive symptoms in intellectual disability: does gender play a role? *Journal of Intellectual Disability Research* **47,** 417–28.

Lunsky Y. & Benson B. A. (1997) Reliability of ratings of consumers with mental retardation and their staff on multiple measures of social support. *American Journal on Mental Retardation* **102,** 280–4.

Lunsky Y. & Benson B. A. (2001) Association between perceived social support and strain, and positive and negative outcomes for adults with mild intellectual disability. *Journal of Intellectual Disability Research* **45,** 106–14.

Lunsky Y., Emery C. F. & Benson B. A. (2002) Staff and self-reports of health behaviours, somatic complaints, and medications among adults with mild intellectual disability. *Journal of Intellectual and Developmental Disability* **27,** 125–35.

Masi G., Brovedani P., Mucci M. & Favilla L. (2002) Assessment of anxiety and depression in adolescents with mental retardation. *Child Psychiatry and Human Development* **32,** 227–37.

Matell M. S. & Jacoby J. (1971) Is there an optimal number of alternatives for Likert Scale Items? Study 1: reliability and validity. *Educational and Psychological Measurement* **31,** 657–74.

Mindham J. & Espie C. A. (2003) Glasgow Anxiety Scale for people with an Intellectual Disability (GAS-ID): development and psychometric properties of a new measure for use with people with mild intellectual disabilities. *Journal of Intellectual Disability Research* **47,** 122–39.

Nanna M. J. & Sawilowsky S. S. (1998) Analysis of Likert scale data in disability and medical rehabilitation research. *Psychological Methods* **3,** 55–67.

Nezu C. M., Nezu A. M., Rothenberg J. L., DelliCarpini L. & Groag I. (1995) Depression in adults with mild

mental retardation: are cognitive variables involved? *Cognitive Therapy and Research* **19,** 227–39.

Novaco R. W. & Taylor J. L. (2004) Assessment of anger and aggression in male offenders with developmental disabilities. *Psychological Assessment* **16,** 42–50.

Payne R. & Jahoda A. (2004) The Glasgow Social Self-Efficacy Scale – a new scale for measuring social self-efficacy in people with intellectual disabilities. *Clinical Psychology and Psychotherapy* **11,** 265–74.

Peter J. P. (1979) Reliability a review of psychometric basics and recent marketing practices. *Journal of Marketing Research* **16,** 6–17.

Powell R. (2003) Psychometric properties of the Beck Depression Inventory and Zung Self Rating Depression Scale in adults with mental retardation. *Mental Retardation* **41,** 88–95.

Prout H. T. & Schaefer B. M. (1985) Self-reports of depression by community-based mildly mentally retarded adults. *American Journal of Mental Deficiency* **90,** 220–2.

Reiss S. & Benson B. A. (1984) Awareness of negative social conditions among mentally retarded, emotionally disturbed outpatients. *American Journal of Psychiatry* **141,** 88–90.

Reynolds W. M. (1979) The utility of multiple-choice test formats with mildly retarded adolescents. *Educational and Psychological Measurement* **39,** 325–31.

Reynolds W. M. & Baker J. A. (1988) Assessment of depression in persons with mental retardation. *American Journal on Mental Retardation* **93,** 93–103.

Reynolds W. M. & Miller K. L. (1985) Depression and learned helplessness in mentally retarded adolescents: an initial investigation. *Applied Research in Mental Retardation* **6,** 295–306.

Robinson J. P., Shaver P. R. & Wrightsman L. S. (1991) Criteria for scale selection and evaluation. In: *Measures of Personality Social Psychological Attitudes* (eds J. P. Robinson, P. R. Shaver & L. S. Wrightsman), pp. 1–15. Academic Press, Inc., San Diego, CA.

Rojahn J., Warren V. J. & Ohringer S. (1994) A comparison of assessment methods for depression in mental retardation. *Journal of Autism and Developmental Disorders* **24,** 305–14.

Russell C. J. & Bobko P. (1992) Moderated regression analysis and Likert scales: too coarse for comfort. *Journal of Applied Psychology* **77,** 336–42.

Schalock R. L., Bonham G. S. & Marchand C. B. (2000) Consumer based quality of life assessment: a path model of perceived satisfaction. *Evaluation and Program Planning* **23,** 77–87.

Schalock R. L., Brown I., Brown R., Cummins R. A., Felce D., Matikka K. D., Keith K. D. & Parmenter T. (2002) Conceptualizations, measurement, and application of quality of life for persons with intellectual disabilities: report of an international panel of experts. *Mental Retardation* **40,** 457–70.

Schloss P. J. (1982) Verbal interaction patterns of depressed and non-depressed institutionalized mental retarded adults. *Applied Research in Mental Retardation* **3,** 1–12.

Senatore V., Matson J. L. & Kazdin A. E. (1985) An inventory to assess psychopathology of mentally retarded adults. *American Journal of Mental Deficiency* **5,** 459–66.

Sigelman C. K. & Budd E. C. (1986) Pictures as an aid in questioning mentally retarded persons. *Rehabilitation Counseling Bulletin* **29,** 173–81.

Sigelman C., Schoenrock C., Spanhel C., Hromas S., Winer J., Budd E. & Martin P. W. (1980) Surveying mentally retarded persons: responsiveness and response validity in three samples. *American Journal of Mental Deficiency* **84,** 479–84.

Sigelman C. K., Budd E. C., Spanhel C. L. & Schoenrock C. J. (1981) Asking questions of retarded persons: a comparison of yes–no and either–or formats. *Applied Research in Mental Retardation* **2,** 347–57.

Sigelman C. K., Budd E. C., Winer J. L., Schoenrock C. J. & Martin P. W. (1982a) Evaluating alternative techniques of questioning mentally retarded persons. *American Journal of Mental Deficiency* **86,** 511–18.

Sigelman C. K., Winer J. L. & Schoenrock C. J. (1982b) The responsiveness of mentally retarded persons to questions. *Education and Training of the Mentally Retarded* **17,** 120–4.

Sigelman C. K., Schoenrock C. J., Budd E. C., Winer J. W., Spanhel C. L., Martin P. W., Hromas S. & Bensberg G. J. (1983) *Communicating with Mentally Retarded Persons: Asking Questions and Getting Answers.* Texas Tech University: Research and Training Center in Mental Retardation, Lubbock, TX.

Stancliffe R. J. (1995) Assessing opportunities for choice-making: a comparison of self- and staff reports. *American Journal on Mental Retardation* **99,** 418–29.

Stancliffe R. J. (2000) Proxy respondents and quality of life. *Evaluation and Program Planning* **23,** 89–93.

Stone A. A., Turkkan J. S., Bachrach C. A., Jobe J. B., Kurtzman H. S. & Cain V. S. (eds) (2000) *The Science of Self-Report. Implications for Research and Practice.* Erlbaum, Mahhwah, NJ.

Verri A., Cummins R. A., Petito F., Vallero E., Monteath S., Gerosa E. & Nappi G. (1999) An Italian-Australian comparison of quality of life among people with intellectual disability living in the community. *Journal of Intellectual Disability Research* **43,** 513–22.