

The Use of Likert Scales With Children

David Mellor,¹ PhD, and Kathleen A. Moore,² PhD

¹Deakin University and ²Charles Darwin University

All correspondence concerning this article should be addressed to David Mellor, School of Psychology, Deakin University, Burwood 3125, Victoria, Australia. E-mail: mellor@deakin.edu.au

Received February 7, 2013; revisions received September 17, 2013; accepted September 24, 2013

Objective We investigated elementary school children's ability to use a variety of Likert response formats to respond to concrete and abstract items. **Methods** 111 children, aged 6–13 years, responded to 2 physical tasks that required them to make objectively verifiable judgments, using a 5-point response format. Then, using 25 items, we ascertained the consistency between responses using a “gold standard” yes/no format and responses using 5-point Likert formats including numeric values, as well as word-based frequencies, similarities to self, and agreeability. **Results** All groups responded similarly to the physical tasks. For the 25 items, the use of numbers to signify agreement yielded low concordance with the yes/no answer format across age-groups. Formats based on words provided higher, but not perfect, concordance for all groups. **Conclusions** Researchers and clinicians need to be aware of the limited understanding that children have of Likert response formats.

Key words abstract tasks; children; Likert scale.

The use of Likert scales, which call for a graded response to a series of statements, is a common means of assessing people's attitudes, values, internal states, and judgments about their own or others' behaviors in both research and clinical practice. Users include professionals such as pediatric psychologists and other health professionals who administer psychometric tests that use Likert scale formats in their research and their practice with children with medical conditions. Since first described by Likert (1932), the range of variables assessed by these scales, as well as their scalar ranges, has proliferated. Further, the populations with which they have been used have expanded to include children as well as adults. However, the degree to which such scalar formats yield valid data when used with children has not been well established. The aim in this article was to investigate this issue.

In his seminal article describing this response format, Likert (1932) reported highly satisfactory reliability data, which he claimed compared favorably with that obtained by other means. Research using adult participants since then has typically confirmed that Likert format scales are generally reliable and valid instruments for the

measurement of a range of attitudes and mood states. They have also been found to yield data that approximate the probability density function thought to fit the data in question even if skewed (e.g., subjective well-being, Cummins, 1998). This distribution is particularly important when measuring, for example, attitudes to stark issues, which require respondents to either agree or disagree. Overall, Likert-type scales provide a useful and relatively simple method of obtaining data in the social sciences.

Recently, Likert scales have been used in a range of research projects and clinical settings in which children are the focus of study or treatment. Some examples of the different scales using Likert formats in research with children, including ages of the samples and response formats used, are presented in Table I. Response scales typically vary from 3 to 5 response points. For example, the Children's Impact of Traumatic Events Scale-Revised (Wolfe, 1996) is based on a 3-point response scale (*very true; somewhat true; not true*), as is the Strengths and Difficulties Questionnaire (Goodman, 1997; *not true; somewhat true; certainly true*). The Social Anxiety Scale for Children-Revised is based on a 5-point scale, with items

Table 1. *Examples of Likert Scales Used in Research and Clinical Practice With Children*

Study/Scale author	Scale	Construct measured	Age range (years)	Response format
Goodman, 1997	Self-report version of the Strengths and Difficulties Questionnaire	General psychological and behavioral problems	4–16	3-point Likert
Harter, 1985	Self-Perception Profile for Children	Self-esteem	8–15	4-point Likert
Kovacs, 1992	Children's Depression Inventory	Depression	7–16	3-point Likert
La Greca & Stone, 1993	Social Anxiety Scale for Children-Revised	Social anxiety	9–13	5-point Likert
McCabe & Ricciardelli, 2002	Children's version of the Eating Attitude Test	Body image concerns, engagement in body change strategies	8–11	5-point Likert
Mellor & Moore, 2003	Questionnaire on Teacher Interaction	Perceived teacher style	11–14	5-point Likert
Moore & Mellor, 2003	Social Interaction Questionnaire	Social anxiety and peer relations	11–14	4-point Likert
Piers & Harris, 1969	Piers-Harris Children's Self-Concept Scale	Self-esteem	7–18	Yes/no
Reynolds & Richmond, 1985	Reynolds Children's Manifest Anxiety Scale	Anxiety	6–19	Yes/no
Wolfe, 1996	Children's Impact of Traumatic Events Scale-revised	Posttraumatic stress	8–16	3-point Likert
Reynolds & Kamphaus, 2004	Behavior Assessment System for Children	Behavioral problems	8–11 and 12–21	Combination of true/false and 4-point frequency
Valla, Bergeron Berube, Gaudet, & St-Georges, 1994	Dominic-R and Terry questionnaires	DSM mental disorders	6–11	Yes/no

rated in terms of how much the item is “true” for the respondent (1 = *not at all*, 5 = *all the time*). Variations include dichotomous choices, for example, “Yes” or “No” responses to items about feelings or behavior (e.g., the Children's Manifest Anxiety Scale and the Piers-Harris Children's Self-Concept Scale), or the selection of one of three statements that best describe the respondent's feelings over the past 2 weeks (Children's Depression Inventory). Responses to the Self-Perception Profile for Children are more complex in that it requires respondents to read two statements, choose the description that best fits them, and then choose whether the description is *really true of them* or *sort of true of them*.

In consideration of the capacity of children to respond to such scales, some authors have been careful in choosing item wording (e.g., Piers-Harris Children's Self-Concept Scale) where items are written at a second-grade reading level, or they have reduced the number of response choices, for example, Wright and Asmundson (2003) who changed the original 5-point Likert scale response format for the Illness Attitudes Scale to a 3-point format to make it more easily understood by children. Other authors have followed Tischer and Lang (1983) and substituted faces on which various degrees of happiness or sadness are depicted for written choice points (e.g., Mellor, McCabe, Ricciardelli, & Ball, 2004).

Despite these variations, little other consideration seems to have been given to the more fundamental issue of whether children actually have the capacity to respond to Likert scale formats in a way that accurately reflects their judgments, attitudes, or values. Cognitive development literature would suggest that this matter is of critical importance. For example, Gelman and Baillargeon (1983) argued that younger children primarily think dichotomously. Thus, asking them to respond on a 5-point scale may be beyond their capacity. With regard to content, Marsh (1986) examined a sample of children aged between 7 and 12 years and found that some children, specifically younger children and those with poor verbal skills, were less able to respond to negatively worded items. Other researchers have tested children aged 5–12 years (Chambers & Craig, 1998; Chambers & Johnstone, 2002) and 5–11 years (von Baeyer, Carlson, & Webb, 1997) and suggested that younger children have a tendency to endorse responses at the extreme end of scales when presented with items based on a Likert scale, thus providing unrefined measures of the constructs under investigation. However, Chambers and Johnston (2002) did suggest that this may vary according to what is being assessed.

The importance of these findings is that, as described above, many scales administered to children are used to

assess intangible theoretical constructs (including emotions) or subjective judgments about the self. These are different from judgments about matters having an objective accuracy (e.g., a number of objects, or people). In Chambers and Johnston's (2002) study, younger children were found to respond as accurately as older children to tasks involving judgments about physical objects, but used extremes in responding to questions about feelings. This pattern was found with both 3-point and 5-point response scales, suggesting that simplifying the response format did not increase children's capacity to use scales.

For a scale to produce reliable and valid data, it must accurately and consistently reflect the measured judgment, attitude, or value. Of critical importance to the use of Likert scales with children is whether an accurate or appropriate internal response will be elicited by the declarative statements. The use of the Likert format assumes that the accurate and representative response has already been internally generated by the child, which may not necessarily be the case. Zeman, Cassano, Perry-Parrish, and Stegall (2006) noted that children's emotional development shares a transactional relationship with their social, neurophysiological, cognitive, and language development. Thus, any scale that uses a Likert format to assess feeling states may be confronted with issues of whether the states are differentiated internally by the child, as well as their cognitive capacity. The work of cognitive developmentalists such as Piaget (1954) would suggest that certain types of judgments should be harder for children in the stage of concrete operations (7–11 years of age), during which the child develops the capacity to make judgments and reason about the physical world, than the subsequent stage of formal operations (11–16 years) in which the capacity to think in abstract terms (usually) evolves. Thus, it would seem that the use of Likert scales for assessing judgments about tangible/physical materials or their representations may be more amenable to assessment in younger children than those about intangible/abstract concepts such as internal feelings. Furthermore, theorists focusing on working memory capacity (e.g., Barrouillet & Lepine, 2005) and on basic arithmetic proficiency (e.g., Haverty, Koedinger, Klahr, & Alibali, 2000) typically support such an age progression in abilities. However, others have found a U-shape exists across ages 7–11 years on mathematical equivalence. For instance, McNeil (2007) found a decrease in performance between the ages of 7 and 9 years, which was reversed by age 11. Of course, children's metacognitive development can also be enhanced and perhaps earlier than the formal operations stage as shown by White and Frederiksen (2005) in their manipulation of metacognitive abilities among fifth-grade children.

This study explored this issue by investigating children's responses to Likert scale items requiring judgments about both physical and abstract concepts. If children are unable to respond accurately to the objectively verifiable and manipulated physical events, then it could be argued that the Likert format cannot accurately assess their judgment about subjective and more abstract matters. On the other hand, if they can respond with accuracy to questions about physical matters, it might be argued that they could have the capacity to use Likert scales in other realms. In line with the findings of Chambers and Johnston (2002), we expected that older children would be able to use Likert formats in both domains, but that younger children's ability would be limited to the concrete physical domain. However, since Zeman, Klimes-Dougan, Cassano, and Adrian (2007) suggested that future research should focus on alternative response formats for assessing children with Likert scales, we also examined a number of alternative anchor points to establish which provides the optimal scale format for all children when abstract constructs are under investigation, in terms of their consistency with a "gold standard" yes/no response. We used yes/no as the gold standard because we believed that it provided the least ambiguity for the participants, as they were not required to respond in terms of degrees of agreement. While Fritzley et al., reported a "yes" bias to this format in a sample of 2–5-year-olds, this bias was found mainly among 2- and 3-year-olds. Other recent research by Rocha, Marche, and Briere (2013) supported the use of the yes/no format in older children. They argued that according to fuzzy trace theory, some forms of multiple-choice questions should elicit higher error rates than yes/no questions.

The alternative anchor formats were numeric values (1–5), as well as word-based frequencies (e.g., *never to regularly*), similarities to self (e.g., *not like me at all, to very much like me*), and agreeability (*strongly agree to strongly disagree*). The rationale for selecting these different Likert anchor formats is that they are used commonly in various measures.

Method

Participants

One hundred and eleven Anglo-Australian children aged between 6 and 13 years ($M = 9.64$ years, $SD = 1.82$) participated in the study. There were 59 girls and 52 boys in the sample. All children were students at elementary schools in a regional city in the state of Victoria, Australia, and were tested on two or three occasions, 2 weeks apart. The sample was divided into three age-groups: 6 and 7 years (<8 years), 8 and 9 years

(<10 years), and 10–13 years. There were 28 children (18 girls, 10 boys), 33 children (19 girls, 14 boys), and 50 children (22 girls, 28 boys) in each group, respectively.

Materials

Physical Tasks

Two physical judgment tasks, called “What I see,” were created to assess responses in Likert format to questions relating to tangible objects. The first task involved a transparent cylindrical beaker with five equidistant vertical points marked on the outside. This grading allowed the beaker to be filled with colored liquid to predetermined levels from five-fifths to one-fifth. The children were then asked to look at three beakers in turn: Two-fifths full, one-fifth full, and four-fifths full of colored liquid. They were asked to mark on a 5-point Likert scale whether the beaker was full, quite full, half full, a little bit full, or not very full. Scores were derived by allocating one point for an answer in the correct direction (relative to the midpoint) for each of the three tasks, and summing these (range: 1–3) owing to the possible lack of precision in the response options “quite full” and “a little bit full.”

The second task involved the child being presented with a reference bowl containing two identical sweets (candy). They were then shown three bowls sequentially that contained three, one, and five sweets, respectively, and asked, “Compared to these bowls of sweets, how many sweets do you have?” The response format was *lots, quite a lot, some, a few, and hardly any*. One point was allocated to responses that were in the correct direction, and the scores summed (range: 1–3).

We chose this lenient approach to scoring (i.e., calling an answer correct if it was in the right direction) to allow for any ambiguity that may have been inherent in the task or instructions.

Abstract Tasks

A 25-item questionnaire was compiled using questions taken from scales used in previous research projects. The items were chosen to cover a range of constructs rather than a single construct, and that would most likely vary in their level of abstractness. Further, they were chosen on the basis of being amenable for wording that would be consistent with the various scale anchor formats. The constructs include body satisfaction, peer relations, adjustment, and perceptions of teacher interactions. Examples of questions included “I sleep very well,” “I feel lonely,” “Spiders make me frightened,” “I feel that parents listen to what their children have to say,” and “I am able to do things as well as most other kids.” All questions are shown in Table II. We also asked a group of 22 elementary

Table II. *Items Used in the Questionnaire and Complexity Rating by Teachers (n = 22)*

Statement	Teacher rating
I sleep very well	Concrete
Meeting new kids makes me frightened	Abstract
I feel that when good things happen they happen because I try hard	Abstract
I enjoy my food	Concrete
My best friend teases me because she/he thinks I am too thin	Abstract
I feel lonely	Concrete
I feel that I am a nice person	Abstract
Being alone makes me frightened	Concrete
On the whole, I like myself	Abstract
I have nightmares	Concrete
I think that cheering more than luck helps a team to win	Abstract
I have lots of energy	Concrete
I feel useless at times	Abstract
I am able to do things as well as most other kids	Abstract
I feel like crying	Abstract
I change my eating to change my body size	Abstract
I feel dumb	Abstract
I feel that parents listen to what their children have to say	Abstract
I feel that whether or not others like you depends on your behavior	Abstract
I believe that some kids are just born good at sports	Concrete
I would like to be top in my class	Abstract
I get anxious doing things when people are watching	Abstract
Spiders make me frightened	Abstract
My mother gives me as much freedom as I want	Abstract
My friends and I have many things in common	Abstract

school teachers to rate the items in the questionnaire as either concrete or abstract because we wanted to understand better the response patterns. While the objective task using the sweets described above was considered to be concrete in nature, it was based on a physical observation and judgment task of a different quality to the language-based tasks in the proposed question sets. The teachers agreed that 7 items were concrete in nature and 18 were abstract (Table II).

Five versions of the questionnaire, each containing the same items but using a different response format, were created. The response formats, following the stem *this statement is like me*, were

1. Yes/no (Scale A).
2. Five points marked by the numbers 1–5 with a key indicating that 1 represented not like me at all, to 5 very much like me (Scale B).

3. Five points marked by the words from strongly agree to strongly disagree (Scale C).
4. Five points marked by the words from never to regularly (Scale D).
5. Five points marked by the numbers 1–5 with a key indicating that 1 represented strong agreement, to 5 representing strong disagreement (Scale E).

Given that Chambers and Johnston (2002) suggested that it is not possible to know the “true” answers to questions about subjective states, and given that it has been proposed that younger children think dichotomously (Gelman & Baillargeon, 1983), the yes/no format was included as the “gold standard” with which to compare the other response formats.

We used 5-point response formats because first, Chambers and Johnston’s (2002) findings showed that children respond in a similar manner to 3- and 5-point scales, and second, Lissitz and Green’s (1975) showed that having more than five points in the response scale provides little by way of extra utility.

Procedure

Permission to conduct the study in government schools was obtained from the State Education authority, and clearance to proceed with the project was obtained from Deakin University Ethics Committee. Principals of five elementary schools were then approached and four agreed to allow information packages to be distributed to parents. The children of those parents who provided signed permission (response rate was 32%) were invited to the testing room in small groups ranging from five for the older children to two for the younger ones. No child whose parents provided consent refused the invitation to participate in the study.

The sessions began with conversations to relax the participants. They were then told that the researchers were interested in how children answer questions, and that they would be asked the same questions several

times over two or three sessions, but on each occasion there would be a different type of answer format. They were shown examples in a suitably worded plain language statement, with differences in the response format pointed out. The students who agreed to continue then signed an assent form.

The first abstract task questionnaire (yes/no format) was then presented, with care taken to ensure the students filled in their demographic details accurately. They were then asked to complete the questionnaire, and to ask for help if they had any difficulties with words, phrases, or questions. They were reminded of this assistance several times. The questions were read to the Year 2 students.

The “objective” physical tasks were completed next, using the questionnaires requiring participants to choose from words referring to relative amounts of colored water in three separate beakers, and a number of sweets in three separate bowls, relative to the standard stimuli. The students were introduced first to the idea of comparison and then asked to make their judgments.

The remaining four questionnaires which used the Likert response formats were administered to the two older groups in a second session, while the youngest group was allowed a third session to complete all questionnaires.

Results

Table III shows the distribution of responses to the two concrete tasks across the three age-groups. In all cases, task by age-group, the majority of children responded in the correct manner on each trial, and obtained a total score of 3. Chi-square analysis for the frequency of responses for each task revealed no association between age-group and response pattern. This finding suggests that compared with older children, younger children responded to the concrete tasks in a similar way.

The analyses of questionnaire data involved comparisons of the responses within each age-group when they

Table III. Percentages of Children in Each Age-Group (%) Achieving Each Score on Physical Tasks^a

Task Score	Task 1: liquid				Task 2: sweets			
	1 (%)	2 (%)	3 (%)	Total	1 (%)	2 (%)	3 (%)	Total
Age-group								
1. 6–8 years, N = 28	11	7	82	100	1	21	78	100
2. 8–10 years, N = 33	6	6	88	100	21	3	76	100
3. 10+ years, N = 50	4	2	94	100	18	10	72	100
Total, N = 111	6	5	89	100	17	11	72	100
	χ^2 (4) 2.89, $p = .58$				χ^2 (4) = 5.95, $p = .20$			

Note. ^aHigher scores (3) indicate correct response on each of the three trials.

used yes/no to their responses and when they used each of the Likert formats. Following the “lenient” approach to scoring (i.e., calling an answer correct if it was in the right direction) used in the analysis of the responses to the concrete tasks, each response on the 5-point scales was recoded into a dichotomous format. The manner of doing this for each scale is summarized in Table IV. This enabled us, on an item by item basis, to ascertain the degree to which the 5-point scale response format answers were consistent with the “gold standard” yes/no response format answers. That is to say, if a child responded “yes” to the question “I feel lonely,” we would expect him/her to have responded with “Agree” or “Strongly agree” when he/she used Scale C, and so on.

We then calculated, for each age-group, the proportions of items for which collapsed responses were

consistent with the response to the yes/no format for each scale format. The results for the teacher-rated Concrete items and the teacher-rated Abstract items are shown separately in Tables V and VI, respectively. We then conducted z tests to compare the response proportions across age-groups. These findings are also included in Tables V and VI.

Teacher-Rated Concrete Items

As can be seen in Tables V and VI, across groups, there was poor consistency between responses using the yes/no format and the responses using Scales B (<35%) and E (<15%), both of which are based on numbers, for the items designated by teachers as Concrete. Using Scale B, the older group (34.78%) exhibited significantly higher concordance than the youngest group (26.26%, $p < .05$,

Table IV. Recoding of Scale Response Formats

Scale format	Response format	Recoded to 0	Recoded to 1	Recoded to 3
A	Yes–No	No	Yes	
B	Numbers 1–5 representing not like me to very much like me	1 = Not like me	2 = A little like me 3 = Like me 4 = A lot like me 5 = Very much like me	
C	Words: Strongly agree–strongly disagree	Disagree Strongly disagree	Strongly agree Agree	Uncertain
D	Words: Never–regularly	Never	Rarely Sometimes Often Regularly	
E	Numbers 1–5 representing strongly agree to strongly disagree	5 = Strongly disagree 4 = Disagree	1 = Strongly agree 2 = Agree	Uncertain

Table V. Concordance of Valid Responses With Yes/No Responses for Each Scale Format by Age-Group for Items Designated as “Concrete” by School Teachers, and Tests for Differences in Proportions Between Groups

Group/comparison	Scale format B (numbers 1–5 with a key indicating that 1 represented <i>not like me at all</i> , through to 5 <i>very much like me</i>)		Scale format C (five points marked by the words from <i>strongly agree</i> to <i>strongly disagree</i>)		Scale format D (five points marked by the words from <i>never</i> to <i>regularly</i>)		Scale format E (five points marked by the numbers 1–5 with a key indicating that 1 represented <i>strong agreement</i> , through to 5 representing <i>strong disagreement</i>)	
	Number congruent	% congruent	Number congruent	% congruent	Number congruent	% congruent	Number congruent	% congruent
Group 1: 6–8 years, $N = 28$	47/179	26.26	127/182	69.78	113/179	63.13	23/171	13.45
Group 2: 8–10 years, $N = 33$	67/224	29.91	177/226	78.32	143/224	63.84	31/224	13.84
Group 3: 10+ years, $N = 50$	121/348	34.78	265/341	77.71	249/336	74.11	30/340	8.82
Group 1 vs. Group 2	z	0.62		1.85*		0.04		0.04
Group 1 vs. Group 3	z	1.89*		1.89*		2.50**		1.47
Group 2 vs. Group 3	z	1.12		0.07		2.50**		1.74*

Note. * $p < .05$, one-tailed test; ** $p < .01$, one-tailed test.

Table VI. Concordance of Valid Responses With Yes/No Responses for Each Scale Format by Age-Group for Items Designated as "Abstract" by School Teachers, and Tests for Differences in Proportions Between Groups

Group/comparison	Scale format B (numbers 1–5 with a key indicating that 1 represented <i>not like me at all</i> , through to 5 <i>very much like me</i>)		Scale format C (five points marked by the words from <i>strongly agree</i> to <i>strongly disagree</i>)		Scale format D (five points marked by the words from <i>never</i> to <i>regularly</i>)		Scale format E (five points marked by the numbers 1–5 with a key indicating that 1 represented <i>strong agreement</i> , through to 5 representing <i>strong disagreement</i>)	
	Number congruent	% congruent	Number congruent	% congruent	Number congruent	% congruent	Number congruent	% congruent
Group 1: 6–8 years, N = 28	150/466	32.19	323/465	69.46	322/463	69.55	90/444	20.27
Group 2: 8–10 years, N = 33	210/586	35.84	398/573	69.46	402/584	68.84	90/585	18.18
Group 3: 10+ years, N = 50	328/891	36.81	642/877	73.20	635/869	73.07	90/880	10.23
Group 1 vs. Group 2	<i>z</i>	1.17		0.07		0.18		1.96*
Group 1 vs. Group 3	<i>z</i>	1.63		1.39		1.30		4.95**
Group 2 vs. Group 3	<i>z</i>	0.32		1.49		1.69		2.86**

Note. * $p < .05$, one-tailed test; ** $p < .01$, one-tailed test.

one-tailed test). Using scale E, the older group (8.82%) did poorer than the middle group (13.45%, $p < .05$, one-tailed test). This was not accounted for by the use of the "Uncertain" midpoint option on Scale E, as the groups did not differ in the proportion of responses that used this option (14.37, 17.26, and 12.96% for youngest, middle, and older groups, respectively).

For Scales C and D, both of which have response formats based on words, agreement on the Concrete items was better (63–78% across groups and scales). Using Scale C, the middle (78.32%) and older groups (77.71%) achieved significantly higher concordance rates than youngest children (69.78%, $p < .01$, one-tailed test, in each case), but did not differ from each other. Using Scale D, the older group (74.11%) achieved higher concordance than both the middle (63.84%) and youngest (63.13%) groups ($p < .01$, one-tailed test, in each case). It should be noted that on Scale C the older group (15.50%) used the "Uncertain" option significantly more as a proportion of their overall responses than did the middle group (7.89%, $z = 2.57$, $p < .01$, one-tailed test).

Teacher-Rated Abstract Items

When the items rated by teachers as Abstract were considered, Scales B and E exhibited similar levels of concordance with the yes/no response format as they did for the Concrete items (32–37% for Scale B, and 10–21% for Scale E across age-groups). Likewise, Scale C (69–72%) and Scale D (69–73%) showed similar, higher levels of concordance with the yes/no format as they did for the Concrete Items. There were no differences in concordance rates across groups, except for Scale E where, as above, the older group (10.23% concordance) did worse than not only the

middle group (18.18%, $p < .01$, one-tailed test), but also the younger group (20.27%, $p < .01$, one-tailed test). In turn, the middle group also achieved lower concordance than the younger group ($p < .05$, one-tailed test). When the use of the "Uncertain" option on scales C and E was investigated, it was found that the middle (21.94%) and older (20.41%) groups used the option in greater proportion of valid responses than the younger group (13.45%) on Scale E ($z = 3.42$ and 3.04 , respectively, $p < .01$, one-tailed tests). Using Scale C, the younger group (10.52% of responses) used the "Uncertain" option significantly less than the middle group (14.53%, $z = 1.84$, $p < .05$, one-tailed test) and the older group (18.13%, $z = 3.59$, $p < .01$, one-tailed test). The older group also used the "Uncertain" option significantly more, proportionately, than the middle group ($z = 1.73$, $p < .05$, one-tailed test).

Discussion

In this study, our aim was to determine whether Likert scale response formats, commonly used to collect data for research or clinical purposes by way of self-report questionnaire, are appropriate for use with children from 6 to 13 years of age. To determine children's understanding of basic graded formats, we used two physical tasks relevant to the stage of concrete operations. A further point of reference was the baseline recording of a yes/no response to the questions posed later. The implications of this research are important for pediatric psychologists so that they can best discern tests and assessments, which are likely to give biased versus accurate responses when evaluating children on various phenomena such as mental illness, traumatic recall, or educational assessments.

The analyses of the data by gender within each age-group and by each scale format revealed no gender differences so we were able to combine their data for subsequent analyses. It is clear that when the children in this sample, ranging in age from 6 to 12 years, were asked to make judgments about physical objects based on 5-point response formats, there was no association between age and response pattern. Even the majority of the youngest children in our study understood how to use a graded scale to make judgments about tasks of a concrete nature. This finding is consistent with that of Chambers and Johnston (2002), who reported that regardless of age, children in their study, who were aged between 5 and 12 years, were capable of answering questions about physical tasks using 3-point and 5-point response formats. When they were asked to respond to questions about more abstract concepts such as their own behaviors, bodily states or emotional states, both younger and older children had more difficulty. When we examined responses across the four different response formats for consistency with the dichotomous yes/no responses, it is clear that when using a numbers to represent *not like me* to *very much like me*, or *strongly agree* to *strongly disagree*, participants in all groups were inconsistent.

Surprisingly, using the scale that required the participants to use numbers to indicate their level of agreement, younger children's responses were more in accord with their yes/no response than older children. While this latter finding may be attributable to older children using the midpoint "Uncertain" response on the scale, it may only be so for the items judged by teachers to be more abstract. For these items, older children used "Uncertain" with greater relative frequency than the other groups. This may suggest that they were using greater discriminating capacity in responding, something they were unable to do with the forced choice dichotomous yes/no response format. The same pattern was evident, however, in the items rated by teachers as concrete.

The word-based response formats produced higher levels of concordance with the yes/no format for all age-groups, but even this was <100%. Generally, in line with expectations, the level of concordance for teacher-rated concrete items for the two formats based on words did exhibit significant increases with age. This general pattern that showed younger children having more difficulty than those in the older groups in using the Likert response formats in a manner consistent with their yes/no responses is consistent with theories of development psychology that suggests that children do not develop the capacity to think and reason in abstract terms until somewhere around the age of 11 or 12 years. For example, Piaget's

theory of cognitive development suggests that at this age children develop the capacity for formal operations and abstract reasoning. Gelman and Baillargeon (1983) suggested that younger children are at a more concrete level, that is, they think in a dichotomous fashion. However, using the word-based response formats in responding to teacher-rated "abstract" items, there were no significant differences between groups on concordance rates. This suggests that younger children can use these formats equally as well as older children, although the concordance with the dichotomous yes/no format was only around 70%.

Notably, the older groups tended to use the mid-scale "Uncertain" response option on the word-based scales relatively more often than did the younger group, suggesting greater discretion and perhaps more thoughtfulness than can be demonstrated in the dichotomous yes/no format. They may be avoiding responding in a black and white mode, or avoiding providing a definitive answer. Younger children may also be more likely to avoid the centre point on scales, or feel a greater obligation to give a determined answer, and respond at extremes (Chambers & Craig, 1998; Chambers & Johnstone, 2002; von Baeyer, et al., 1997).

These findings have significant implications for the way clinical or research information is elicited from children <10 years of age. As discussed above, it is not difficult to find examples of research instruments and clinical tests that use Likert type response formats with children. If children even up to age 12 years are not able to use response formats based on numbers in a manner consistent with their dichotomous responses, the information obtained from them will be neither reliable nor valid, assuming that the more basic yes/no format provides a true reflection of their inner state or feelings. With the ever increasing push toward efficiency in clinical services, depending on the weight given to the particular measure being administered, decision-making may be flawed.

Similarly, researchers who are interested in the study of children's development may need to pay attention to the way their research instruments are worded, particularly the response formats. It appears that the proportion of children, regardless of age, responded to the items in a manner consistent with their yes/no response varied from very low (approximately 10%) to moderate (approximately 70%), depending on response format. If this is the case more generally, much of the data elicited from children may be unreliable and invalid. On the positive side, our findings suggest that younger children show a greater capacity to master a 5-point Likert scale when the anchors are worded

in a way that is meaningful to them, however, the percentage of those who do so remains low in comparison with older children.

Despite these warnings, the findings of our study should be viewed as preliminary indications of the need for caution when assessing or researching with children. Our sample was not large, particularly when we divided it into three age-groupings. In addition, the items used in the questionnaires and indeed in the concrete tasks may not represent adequately the range of tasks of either kind. While the questionnaire tasks were more abstract than the concrete task, the degree of abstractness of items used in the questionnaire varied somewhat, partly because the questions were drawn from a number of different questionnaires. For example, questions such as “I change my eating to change my body size” asked the child to reflect on behaviors, while questions such as “I would like to be top in my class” and “I feel lonely” asked the child to reflect on ambitions or subjective states. While we asked teachers to rate our items for their level of abstractness, it is difficult to standardize level of abstractness in a question, and as the current study was exploratory in nature, future studies might attempt to address this issue.

Having children complete five versions of the 25-item questionnaire, even with a different response format each time, and over two or three separate occasions, may still have influenced the outcomes. For example, a potential confound between questionnaire order and the lower reliability of results may have occurred on the version that asked about level of agreement by numbers because its administration was last out of the five different versions of the questionnaire. A larger study with counterbalanced design would be a useful next step. Perhaps a shorter questionnaire would suffice.

Future studies might also explore the optimum number of points on the response scale for children of various ages. We used 5-point response formats on the basis of Lissitz and Green’s (1975) suggestion that having more than 5 points in the response scale provides little by way of extra utility, and Chambers and Johnston’s finding that children responded in a similar manner to 3- and 5-point response formats. However, it would be useful to know whether a 3-point or a 7-point Likert scale is in fact more reliable for younger children when used with different anchors. Nunnally (1967) observed that in terms of psychometric theory, up to seven response options typically increase internal reliability, and beyond that number reliability plateaus. Increasing the number of subscale items also increases internal reliability, but the current study was not concerned with particular constructs or subscales. Cox (1980) also concluded that 7-point

scales may be optimal. However, none of these writers were focusing on the use of scales with children.

Similarly, while our questionnaire was not designed to measure any particular constructs, and all items were positively worded and relatively simple, future studies might explore how negatively worded items that are part of a unidimensional scale are managed by children of different ages. Benson and Hocevar (1985) reported that children in Grades 4–6 had difficulty indicating agreement when required to disagree with a negative statement, and Marsh (1986) reported that younger children, and children with poorer reading skills, are less able to respond appropriately to negative items and that this effect biases the interpretation of their responses.

It would also be of interest to investigate the ability of children to respond to questionnaires using Likert scale response formats, with a variety of anchors when different constructs are being assessed. We used a sample of questions drawn from a range of questionnaires to avoid confounding content with response ability. However, another approach may be to use a number of scales that assess particular constructs (e.g., body satisfaction, peer relations, adjustment, and perception of teacher interactions) to explore whether the reliability of the Likert scales varies with age depending on the construct being assessed.

Conclusion

Likert scales have been used to assess various aspects of children’s feelings, beliefs, and attitudes with little critical evaluation of their validity. Our findings suggest that even though there is little difference in the abilities of older and younger children in their ability to make judgments about tangible objects or abstract questions using response formats based on words, most children might have difficulty providing answers when the Likert response formats are based on numbers. The easiest format for children appears to be that which is based on words that reflect frequency of behaviors/thoughts (i.e., never to regularly). Therefore, the reliability of Likert scales formatted in this way, including those listed in Table I, is likely to be higher than if other formats were used. While further research is required, our findings suggest that pediatric researchers and clinicians should be cautious when using scales with Likert-type response formats and seek collaborating information from a range of sources including parents, teachers and, where relevant, peers. Such collaborating statements are especially important if the child under pediatric assessment has any known or suspected learning or behavioral difficulties.

Funding

Conflicts of interest: None declared.

References

- Barrouillet, P., & Lepine, R. (2005). Working memory and children's use of retrieval to solve addition problems. *Journal of Experimental Child Psychology*, 91, 183–204.
- Benson, J., & Hocevar, D. (1985). The impact of item phrasing on the validity of attitudinal scales for elementary school children. *Journal of Educational Measurement*, 22, 231–240. doi:10.1111/j.1745-3984.1985.tb01061.x
- Chambers, C. T., & Craig, K. D. (1998). An intrusive impact of anchors in children's faces pain scales. *Pain*, 78, 27–37. doi:10.1016/S0304-3959(98)00112-2
- Chambers, C. T., & Johnston, C. (2002). Developmental differences in children's use of rating scales. *Journal of Pediatric Psychology*, 27, 27–36.
- Cox, E. P. (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research*, 27, 407–422. doi:10.2307/3150495
- Cummins, R.A. (1998). The second approximation to an international standard of life satisfaction. *Social Indicators Research*, 43, 307–334. doi:10.1023/A:1006831107052
- Gelman, R., & Baillargeon, R. (1983). A review of some piagetian concepts. In J.H. Flavell, & M. Markman (Eds.), *Handbook of child psychology, Volume 3: Cognitive development*. New York, NY: Wiley.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x
- Harter, S. (1985). *Manual for the self-perception profile for children*. Denver, CO: University of Denver, Department of Developmental Psychology.
- Haverty, L. A., Koedinger, K. R., Klahr, D., & Alibali, M. W. (2000). Solving inductive reasoning problems in mathematics: Not so trivial pursuit. *Cognitive Science*, 24, 249–298.
- Kovacs, M. (1992). *Children's Depression Inventory manual*. Toronto, ON, Canada: Multi-Health Systems, Inc.
- La Greca, A. M., & Stone, W. L. (1993). Social anxiety scale for children—revised: Factor structure and concurrent validity. *Journal of Clinical Child Psychology*, 22, 17–27. doi:10.1207/s15374424jccp2201_2
- Likert, R. (1932). A technique for the measurement of attitudes. *Journal of Applied Psychology*, 39, 31–36.
- Lissitz, R.W., & Green, S. B. (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology*, 60, 10–13.
- Marsh, H. W. (1986). Negative item bias in rating scales for pre-adolescent children. *Developmental Psychology*, 22, 37–49. doi:10.1037/0012-1649.22.1.37
- McCabe, M., & Ricciardelli, L. (2002). Body image and strategies to lose weight and increase muscle among boys and girls. *Health Psychology*, 22, 39–46. doi:10.1037/0278-6133.22.1.39
- McNeil, N. M. (2007). U-shaped development in math: 7-year-olds outperform 9-year-olds on equivalence problems. *Developmental Psychology*, 43, 687–695.
- Mellor, D., McCabe, M., Ricciardelli, L., & Ball, K. (2004). Body image importance and body dissatisfaction among Indigenous Australian adolescents. *Body Image*, 1, 289–297. doi:10.1016/j.bodyim.2004.05.003
- Mellor, D., & Moore, K. (2003). The Questionnaire on Teacher Interaction: Assessing information transfer in single and multi-teacher environments. *Journal of Classroom Interaction*, 38, 29–35.
- Moore, K., & Mellor, D. (2003). The nature of children's interactions at school. *School Psychology International*, 24, 329–339. doi:10.1177/01430343030243005
- Nunnally, J. (1967). *Psychometric theory*. New York, NY: McGraw Hill.
- Piaget, J. (1954). *The construction of reality in the child*. New York, NY: Basic Books.
- Piers, E. V., & Harris, D. B. (1969). *The Piers-Harris Children's Self-Concept Scale*. Nashville, TN: Counselor Recordings and Tests.
- Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children-2 (BASC-2)*. Circle Pines, MN: American Guidance Service.
- Reynolds, C. R., & Richmond, B. O. (1985). *Revised Children's Manifest Anxiety Scale (RCMAS) manual*. Los Angeles, CA: Western Psychological Services.
- Rocha, E. M., Marche, T. A., & Briere, J. L. (2013). The effect of forced-choice questions on children's suggestibility: A comparison of multiple-choice and yes/no questions. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement*, 45, 1–11. doi:10.1037/a0028507
- Tischer, M., & Lang, M. (1983). The Children's Depression Scale: Review and further developments. In D. P. Cantwell, & G. A. Carlson (Eds.), *Affective*

- disorders in childhood and adolescence* (pp. 181–203). Jamaica, NY: Spectrum.
- Valla, J. P., Bergeron, L., Berube, H., Gaudet, N., & St-Georges, M. A. (1994). A structured pictorial questionnaire to assess DSM-III-R-based diagnoses in children (6–11 years): Development, validity, and reliability. *Journal of Abnormal Child Psychology*, 22, 403–423. doi:10.1097/00004583-200001000-00020
- von Baeyer, C., Carlson, G., & Webb, L. (1997). Underprediction of pain in children undergoing ear piercing. *Behavioral Research and Therapy*, 35, 399–404. doi:10.1016/S0005-7967(96)00127-1
- White, B., & Frederiksen, J. (2005). A theoretical framework and approach for fostering metacognitive development. *Educational Psychologist*, 40, 211–223.
- Wolfe, V. V. (1996). Measuring post-traumatic stress disorder: The Children's Impact of Traumatic Events Scale-Revised. *The APSAC Advisor*, 9, 25–26.
- Wright, K. D., & Asmundson, G. J. G. (2003). Health anxiety in children: Development and psychometric properties of the Childhood Illness Attitude Scales. *Cognitive Behaviour Therapy*, 34, 194–201. doi:10.1080/16506070310014691
- Zeman, J., Cassano, M., Perry-Parrish, C., & Stegall, S. (2006). Emotion regulation in children and adolescents. *Journal of Developmental and Behavioral Pediatrics*, 27, 155–168.
- Zeman, J., Klimes-Dougan, B., Cassano, M., & Adrian, M. (2007). Measurement issues in emotion research with children and adolescents. *Clinical Psychology: Science and Practice*, 14, 377–401. doi:10.1111/j.1468-2850.2007.00098