

Developmental Differences in Children's Use of Rating Scales

Christine T. Chambers, PhD, and *Charlotte Johnston*, PhD
University of British Columbia

Objective: To examine the effect of child age and number of response choices on children's tendency to respond at the extremes of Likert-type scales rating emotional states.

Methods: Sixty children (5–6 years, 7–9 years, 10–12 years) were randomly assigned to use either three or five response choices in providing ratings in three different task conditions. Tasks were designed to have correct choices at the midpoints of the rating scales. Children also completed a self-report feelings questionnaire.

Results: Results showed that younger children responded in an extreme manner when rating emotion-based, but not physical, tasks. Children's extreme scores did not vary as a function of number of response choices used. More extreme scores on the three tasks were related to more extreme scores on the feelings questionnaire.

Conclusions: These results indicate that young children may respond in an extreme manner when rating emotional states. Researchers and clinicians should take this into account when interpreting children's self-report ratings.

Key words: *children; assessment; methodology; self-report; rating scales.*

Children often must provide self-reports of subjective, emotional states (La Greca, 1990). Over the years, many questionnaires assessing a variety of psychological content areas (e.g., depression, anxiety) have been developed to elicit reports from children (for a review, see Rodrigue, Geffken, & Streisand, 2000). When responding to these questionnaires, children are typically asked to rate, on a Likert-type scale, the level at which they agree with statements or the degree to which certain items apply to them. Although the specific Likert-type scale used tends to vary, ranging from only two re-

sponse options up to seven, scales typically ask children to choose from among a series of graded response choices (e.g., "not at all," "a little," "somewhat," "quite a bit," or "a lot"). In pediatric psychology research, children's responses on these self-report questionnaires are used to examine factors related to adjustment and psychopathology, as well as to evaluate treatment outcome. In clinical practice, children's responses to these scales are often used to make diagnostic and treatment decisions.

The use of children's ratings in drawing empirical research conclusions and making clinical decisions assumes that the Likert-type rating scales are appropriate for use by children and that the ratings they produce are valid. However, one might expect that, due to their limited cognitive abilities, some

All correspondence should be sent to Christine T. Chambers, Centre for Community Health and Health Evaluation Research, 4480 Oak Street, L408, Vancouver, British Columbia, V6H 3V4, Canada. E-mail: cchambers@cw.bc.ca.

children (particularly younger children) might experience difficulties using these scales. Whereas there has been critical discussion of the implicit and explicit assumptions of issues in using rating scales in developmental research (Surber, 1984), to date, the only study examining such measurement issues among children was conducted by Marsh (1986). This study examined the negative item bias, which is produced when negative items on rating scales require a higher level of verbal reasoning than positive items (e.g., "I am not smart at mathematics" vs. "I am smart at mathematics"). Indeed, Marsh found that younger children, and children with poorer reading skills, were less able to respond appropriately to negative items on questionnaires and that this effect biased the interpretation of the children's responses. However, this study examined only the impact of item phrasing on children's responses and did not explore the issue of response choice appropriateness. In fact, although considerable research has examined response styles among adults (for a review, see Paulhus, 1991), this research has not been extended to children. Despite the lack of research specifically examining response styles among children, we have noted in our research, anecdotally, that younger children tend to respond in an extreme manner (i.e., endorsing the options at either end of the response continuum) when asked to use Likert-type rating scales (e.g., Chambers & Craig, 1998). For example, younger children seem to be more likely to respond that they are either "a lot" or "not at all" happy, whereas older children seem more capable of providing graded ratings in the middle of the scale (e.g., "a little" happy). This observation has also been made by other researchers (e.g., Goodenough et al., 1997; von Baeyer, Carlson, & Webb, 1997).

Research from a developmental perspective supports the observation that the use of Likert-type rating scales would be more difficult for younger than older children. According to Piagetian theory, young children characteristically engage in dichotomous thinking (Gelman & Baillargeon, 1983) and may therefore focus only on the two extremes of Likert-type rating scales. Furthermore, research examining the ability of school-age children to acknowledge that they can experience multiple feelings at the same time (e.g., feeling both happy and sad) indicates that this is a difficult developmental task (for a review, Harter & Whitesell, 1989). Specifically, there is evidence that children progress

from an inability to acknowledge simultaneous emotions (at the age of approximately 4 years), to acknowledgment of the simultaneity of emotions when they are of the same valence and directed at the same target, and then to acknowledgment of emotions of different valence directed at a different target, and finally progressing to acknowledgment of emotions of different valence directed at the same target (at the age of approximately 12 years) (Harter & Buddin, 1987). One might speculate that asking children to provide fine-grained intensity ratings of feelings (e.g., acknowledging that they feel "a little" happy and "quite a bit" sad) falls along a similar continuum of developmental complexity, with the ability to identify with and report on the full intensity of emotions or feelings preceding the ability to report on more subtle levels of intensity. Therefore, the developmental research examining children's understanding of the simultaneity of emotions is consistent with the prediction that providing Likert-type ratings of emotions would be more difficult for younger children than older children.

Often in psychological research, children are asked to rate subjective, unobservable states, and it is impossible to disentangle the "truth" in children's responses to such questions (i.e., whether extreme responses are truly reflective of inner states or whether this is a particular response style or bias). The purpose of this study was to examine age-related differences in school-age children's use of rating scales when responding to a series of tasks where the investigators established and manipulated the "truth" of the ratings. Three types of tasks were examined, including children's ratings of physical characteristics (physical task), other people's feelings (social objective task), and the children's own feelings (subjective task). These three task types were chosen as gradations in the degree to which the "true" responses were known. For the physical task, the true and correct answer was obvious and objective, whereas for the subjective task, the correct answer is somewhat more subjective. However, even for this task, with the information given, it is reasonable to assume that most children would provide the experimenter-anticipated correct response. Visual cues were provided in the physical task; however, the social objective and subjective tasks were presented without a visual accompaniment in order to most closely approximate the more abstract task of reporting on subjective feel-

ings and to resemble commonly used children's self-report measures. As the ability to accurately use rating scales might differ due to the number of options included in the response array, a second purpose of this study was to examine differences in children's ratings using two different numbers of response options (i.e., three choices or five choices). In addition, this study examined the relationship between children's tendencies to respond at the extremes of rating scales (i.e., endorsing either "not at all" or "a lot") in the manipulated tasks and their responses on a self-report feelings questionnaire.

We hypothesized that younger children would have more extreme rating scale scores across tasks, and on the self-report feelings questionnaire, than older children. We also expected that children using the five response options (which require more fine-grained discriminations of emotions) would have more extreme scores across tasks than children using only three response options. Finally, we expected that the tendency to respond at the extremes would be significantly related across tasks and that more extreme scores on the manipulated tasks would be related to more extreme scores on the subjective, feelings questionnaire.

Method

Participants

Participants were 60 children (31 girls and 29 boys) between the ages of 5 and 12 years ($M = 8.07$, $SD = 2.37$), recruited from advertisements placed in local libraries, community newspapers, and children's groups. Children were stratified into three age categories: 5 to 6 years, $n = 20$ (10 girls, 10 boys); 7 to 9 years, $n = 20$ (12 girls, 8 boys); 10 to 12 years, $n = 20$ (9 girls, 11 boys) to be consistent with prior research examining age-related differences in children's understanding of emotions (Banerjee, 1997). They were randomly assigned either to a three-choice ($n = 30$; 18 girls, 12 boys) or five-choice ($n = 30$; 13 girls, 17 boys) response group. Assignment to the three- and five-choice response groups was balanced within age groups. Children came from middle to upper class families ($M = 24.96$, $SD = 11.06$; Hollingshead Index Level II; Miller, 1977). The University of British Columbia Behavioural Research Ethics Committee approved this study. Written informed consent and basic demographic

information was obtained from parents, and verbal assent was obtained from children. Children received a certificate and sticker to acknowledge their participation.

Procedure

Children were tested individually in a quiet room in the University of British Columbia Psychology Clinic ($n = 57$) or a quiet area in their own home ($n = 3$) and were tested either by one of us (CTC; $n = 40$) or a trained undergraduate research assistant ($n = 20$). After receiving verbal assent, children were introduced to the response choices they would be using. The experimenter gave the following instructions to the child, "Today I'm going to be asking you some questions about how you feel and how you think other people might look or feel. I will give you some choices to pick your answer from. See here, these are the different choices you can pick your answer from. There aren't any right or wrong answers to the questions, I just want to hear what you think. So, whenever I ask you a question, you can either say 'not at all,' 'somewhat,' or 'a lot' (or if in the five-choice response group, 'not at all,' 'a little,' 'somewhat,' 'quite a bit,' or 'a lot'). Can you please repeat those for me out loud?" Children were shown their response options on a 6 inch \times 12 inch piece of laminated paper. The response options were in 38-point bold font, and each choice appeared in a 2 inch \times 2 inch text box. If the child was unable to repeat the response options on his or her own, the experimenter repeated them until the child was able to do so. Children were then asked to complete each of the three tasks described below, as well as a seven-item feelings questionnaire that asked children to rate how happy, sad, angry, excited, calm/relaxed, nervous/worried, scared/afraid they were feeling during the day they were tested (i.e., "today") (Chambers & Craig, 1998). The tasks and questionnaire were administered verbally to the children. Children within each response choice group used the same response options (three or five) for the three different tasks and the feelings questionnaire. The response options were available to children continuously through the testing procedure and were reviewed with children between tasks. The order in which the children completed the three tasks and feelings questionnaire was randomized for each child, and the order in which individual items within the tasks or questionnaire

Table I. Listing of Test and Filler Item Content for the Physical, Social Objective, and Subjective Tasks

	Physical	Social objective	Subjective
Test items			
1	Children with different numbers of books	Children with different numbers of bags of candy	Inviting ten children to a birthday party and five come
2	Children with different sized gym balls	Children who get different amounts of allowance money	Wanting to see all of a TV show and seeing half
3	Children with different lengths of hair	Children who are sick for different numbers of days	Wanting to go to the movies today and going tomorrow
4	Children with different numbers of balloons	Children whose parents are late picking them up	Sharing cake with one child, and a friend sharing with three
Filler items			
1	Children of different heights	Children winning different positions in a race	Losing a very favorite toy
2	Children wearing different amounts of red clothing	Children getting picked at different positions for a team	Coming in last in a race

Physical tasks were presented as cartoon pictures with accompanying verbal stories. An example of a physical test item is shown in Figure 1. Social objective and subjective tasks were presented as brief verbal stories. Copies of the cartoon pictures and stories are available from the authors upon request.

were presented was also randomized. The testing took approximately 15 to 20 minutes per child.

Tasks

Complete listings of the content of items for the tasks described appear in Table I. Cartoons and/or stories were used to form the basis for these tasks, as they have been frequently found useful as a methodological tool in examining children's understanding of emotions (e.g., Donaldson & Westerman, 1986).

Physical Task. Children were shown six separate sets of cartoon pictures presented on 12 inch \times 12 inch pieces of laminated paper (see Figure 1 for an example). Each picture depicted four children who displayed a different physical characteristic (e.g., carrying different numbers of books: one child with no books, one child with two books, one child with six books, and one child with eight books). A verbal description accompanied the picture (e.g., "Today is library day and the children are returning their library books. Different children have different numbers of books"). After a 5-second delay, the picture was removed and children were shown a second picture that showed only one child (i.e., the target child) who presented the same physical characteristic as the children shown in the previous picture (e.g., carrying four books). Children were then asked, "Here is Mike. How many books does Mike have compared to the other children?" The physical task consisted of four test items (i.e., items in which

the target child should be rated in the middle), as with the example described above and shown in Figure 1, and two filler items (i.e., items in which the target child should be rated at the extremes) (e.g., the target child being the tallest child in comparison to the other four children). Such filler items were included to avoid children concluding that the correct response was always in the middle of the rating scale.

Social Objective Task. Children were told six brief stories about four children whose experience of a situation varied. For example, "It's Halloween and all of the children have just come home from Trick or Treating. These children love to eat candy and the more candy they get, the happier they are. One child got four bags of candy, one child got three bags of candy, one child got one bag of candy, and one child got no bags." Then, children were asked to rate how they thought a target child would feel, "Amanda got two bags of candy. How good would Amanda feel compared to the other children?" This task was presented without a visual accompaniment to most closely approximate the more abstract task of reporting on subjective feelings. The social objective task consisted of four test items (i.e., items in which the target child should be rated in the middle), as with the example described above, and two filler items (i.e., items in which the target child should be rated at the extremes) (e.g., the target child coming in first in a race). Of the four test items, two asked children how "good" they thought the target child would feel, and two asked the chil-

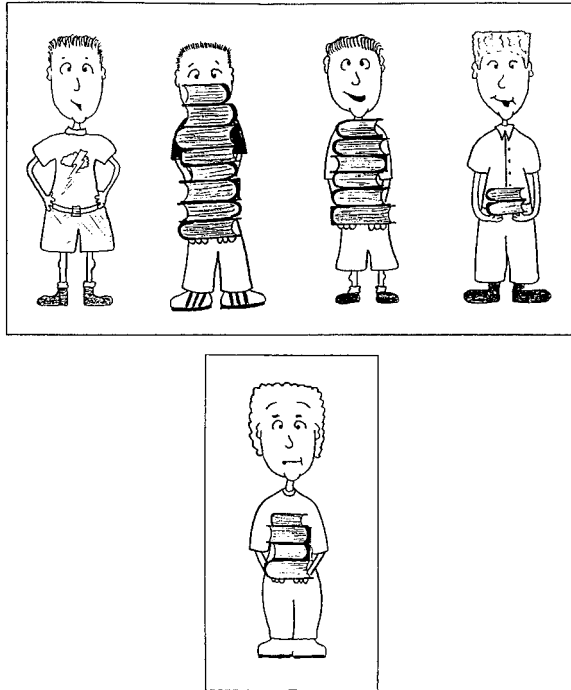


Figure 1. Example of a physical task item. The upper box is an example of the first picture children were shown within an item, with four children varying on a physical characteristic. The lower box is an example of the second picture children were shown, of the target child.

dren how “bad” they thought the target child would feel. Similarly, of the two filler items, one asked children how “good” they thought the target child would feel, and one asked the children how “bad” they thought the target child would feel.

Subjective Task. Children were told six brief stories in which they were asked to imagine themselves being in different situations. For example, “You invite 10 children to your birthday party and 5 of them come.” Then, children were asked to rate, “How good would you feel?” Similar to the social objective tasks, the subjective task consisted of four test items where the correct responses were likely to be in the middle, and two filler items where the correct responses were likely to be at the extremes (e.g., losing their *very favorite* toy). While it is possible that there may have been some individual differences in how the children responded and reacted to the subjective tasks situations, we anticipated that for most children their reactions would match our designated correct responses. This assumption is supported by research showing that by 5 years of age children have a good understanding of situations

that elicit certain emotions (Harris, 1983). Of the four test items, two asked children how “good” they would feel, and two asked the children how “bad” they would feel. Similarly, of the two filler items, one asked children how “good” they would feel, and one asked how “bad” they would feel.

Scoring

Children’s responses using the first (i.e., “not at all”) or last (i.e., “a lot”) options on the three- and five-response choice rating scales were scored as extreme responses for all task and questionnaire items. To reflect the degree with which children responded in an extreme manner, we summed the number of task items to which children responded at the extremes (i.e., endorsed as either “not at all” or “a lot”) to yield extreme scores ranging from 0 to 4 for each task type (higher scores reflect a greater tendency to respond at the extremes). For the feelings questionnaire, we summed the number of items to which children responded at the extremes (i.e., endorsed as either “not at all” or “a lot”) to yield an extreme score ranging from 0 to 7 (higher scores reflect a greater tendency to respond at the extremes).

Results

A 2 (Number of Response Choices: Three vs. Five) \times 3 (Age Group: 5–6 years vs. 7–9 years vs. 10–12 years) \times 3 (Task Type: Physical vs. Social Objective vs. Subjective Task) between-within analysis of variance (ANOVA) was used to examine differences in children’s extreme scores for the test items. (Copies of complete ANOVA tables are available from the authors upon request). Results showed that there were significant main effects for both age group, $F(2, 54) = 9.89, p < .001$, and task type, $F(2, 108) = 35.12, p < .001$, on children’s extreme scores for the test items. Follow-up Student Newman Keuls post-hoc testing revealed that 5- to 6-year-olds had more extreme scores across tasks and number of response choices than both the 7- to 9-year-olds and the 10- to 12-year-olds; there were no differences between 7- to 9-year-olds’ and 10- to 12-year-olds’ extreme scores. Further, the post-hoc testing revealed that children, as a group collapsing across tasks and number of response choices, showed more extreme scores when rating the subjective task than for the social objective task, which in turn elicited more ex-

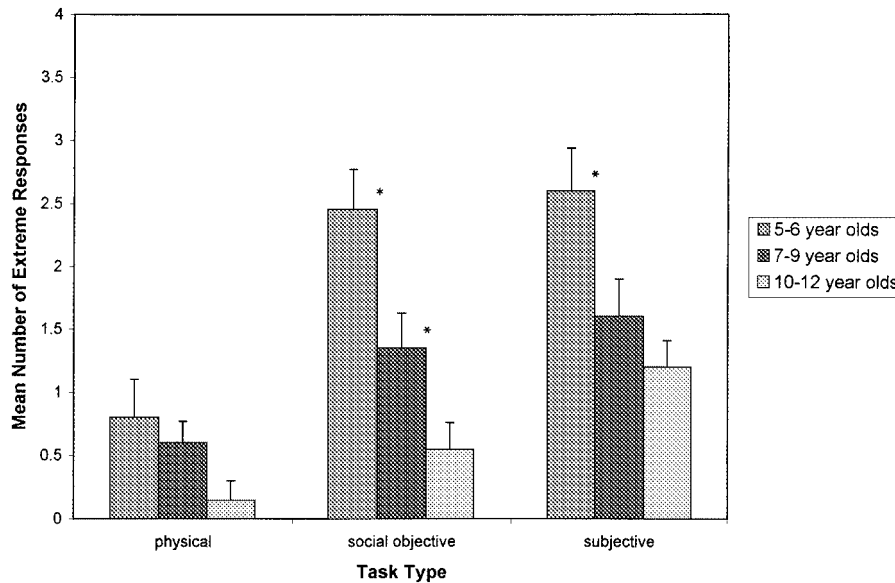


Figure 2. Children's extreme scores on test items as a function of age group and task type. * = significant difference at $p < .05$ or better.

treme scores than the physical task. The main effect of number of response choices was not significant, $F(1, 54) = 1.42, p > .05$.

Further, the results from the ANOVA indicated that the Age Group \times Task Type interaction was significant, $F(4, 108) = 3.04, p < .05$. The Age Group \times Task Type interaction is shown graphically in Figure 2. No other interactions were significant. Three follow-up one-way ANOVAs with Student Newman Keuls post-hoc tests were conducted to examine the simple main effects of age group for each task type. For the physical task, the main effect of age group was not significant, $F(2, 57) = 2.50, p > .05$. For the social objective task, the main effect of age group was significant, $F(2, 57) = 12.00, p < .001$; the post-hoc tests indicated that the 5- to 6-year-olds had significantly more extreme scores than the 7- to 9-year-olds and the 10- to 12-year-olds, and that the 7- to 9-year-olds had significantly more extreme scores than the 10- to 12-year-olds. For the subjective task, the main effect of age group was also significant, $F(2, 57) = 6.25, p < .005$; the post-hoc tests indicated that the 5- to 6-year-olds had significantly more extreme scores than the 7- to 9-year-olds and the 10- to 12-year-olds. In partial support of our hypothesis, younger children tended to respond in an extreme manner in comparison to older children when rating social objective and subjective, but not physical, tasks. In contrast to our hypothesis, the number of response options included in the array (i.e., three versus five) did not have an effect on

children's tendency to respond at the extremes on the three tasks.

A 2 (Number of Response Choices: Three vs. Five) \times 3 (Age Group: 5–6 years vs. 7–9 years vs. 10–12 years) between-subjects ANOVA was used to examine differences in children's extreme scores on the self-report feelings questionnaire. Results showed that the main effect of age group was significant, $F(2, 54) = 9.68, p < .001$. Follow-up Student Newman Keuls post-hoc tests indicated that the 5- to 6-year-olds ($M = 5.95, SD = 1.19$) had significantly more extreme scores than the 7- to 9-year-olds ($M = 4.75, SD = 1.25$) and the 10- to 12-year-olds ($M = 4.00, SD = 1.84$). The main effect of number of response choices, $F(1, 54) = 2.70, p > .05$, and the Age Group \times Number of Response Choices interaction, $F(2, 54) = 3.65, p > .05$, were not significant. Therefore, as hypothesized, younger children tended to respond in an extreme manner compared to the older children when providing ratings on the self-report feelings questionnaire. Again, the number of response choices provided in the array (i.e., three versus five) did not have an effect on children's tendency to respond at the extremes on the self-report feelings questionnaire.

Pearson product moment correlations were used to examine the relationship among children's extreme scores on test items for the three tasks and their extreme scores on the feelings questionnaire. As shown in Table II, all correlations were significant at $p < .05$ or better, indicating that the ten-

Table II. Correlations Between Children's Extreme Scores on Test Items for the Three Tasks and Feelings Questionnaire

	Physical	Social obj.	Subjective	Feelings questionnaire
Physical task	—	.51**	.47**	.27*
Social obj. task		—	.62**	.31*
Subjective task			—	.27*
Feelings questionnaire				—

Higher scores reflect a greater tendency to respond at the extremes.

* $p < .05$.

** $p < .001$.

dency to respond at the extremes was correlated across tasks and was significantly related to children's tendency to respond at the extremes on the self-report feelings questionnaire. Of note, when the effect of age group was controlled for, the correlations among the tasks remained highly significant (ranging from $r = .40$ to $.52$).

To ensure that the pattern of results could be attributed to the tendency of younger children to respond at the extremes of rating scales, and not simply the hesitancy of older children to use the extremes, we also analyzed children's responses on the filler items (where extreme ratings were correct) for each task. We summed the number of items children responded at the extremes on (i.e., endorsing either "not at all" or "a lot") for the filler items for the three tasks to yield extreme scores ranging from 0 to 2 for each task type. A 2 (Number of Response Choices: Three vs. Five) \times 3 (Age Group: 5–6 years vs. 7–9 years vs. 10–12 years) multivariate analysis of variance (MANOVA), with scores on each of the three task types as dependent measures, was conducted to examine differences in children's extreme scores on the filler items for the three tasks. We chose this multivariate approach as there were no specific a priori research questions regarding differences in filler items scores as a function of task type. Results showed that the main effects of age group and number of response choices, and their interaction, were not significant for any of the three tasks. The means (with standard deviations in parentheses) for children's extreme scores on the filler items, collapsing across age and number of response choices, for the physical, social objective, and subjective tasks were 1.38 (0.58), 1.72 (0.58) and 1.50 (0.65), respectively. In other words, most of the children, regardless of age and response choice group, were capable of providing ratings at the extremes of the rating scales when necessary.

Discussion

This study examined age differences in children's use of Likert-type rating scales when responding to a series of tasks where the "truth" of the tasks was established and manipulated by the investigators. As expected, younger children had more extreme scores than older children, but only for the social objective and subjective tasks, not the physical task. The lack of age-related differences in children's ratings for the physical task was due to the fact that most of the children, regardless of age, were able to provide accurate ratings for that task type (i.e., on average, children of all ages provided an extreme response for fewer than one of the four items on the physical task). For the social objective task, 5- to 6-year-olds had more extreme scores than the 7- to 9-year-olds and the 10- to 12-year-olds, and, in turn, the 7- to 9-year-olds had more extreme scores than the 10- to 12-year-olds. Similarly, for the subjective task, the 5- to 6-year-olds had more extreme scores than the 7- to 9-year-olds and the 10- to 12-year-olds. The lack of significant differences between the 7- to 9-year-olds' scores and the 10- to 12-year-olds' scores on the subjective task is likely due to the fact that even the oldest children showed elevated extreme scores for this task type. The corresponding effect size for this interaction was moderate ($\eta^2 = .10$) and the effect sizes for the main effects of task and age were large ($\eta^2 = .56$ and $.27$, respectively) indicating that these findings were relatively robust (Cohen, 1988). As the investigators manipulated the "truth" of the tasks, these results suggest that the accurate use of rating scales is a difficult developmental task for young elementary school-age children.

In addition, as all children, regardless of age, were capable of providing accurate ratings for the *physical* task, but not for either of the two tasks that involved rating emotions, this indicates that there may be something specific about providing ratings of emotions that could account for the tendency of children, particularly younger children, to respond at the extremes. In support, the main effects analyses for task type revealed that as the tasks became more subjective and emotion focused, children's extreme scores, regardless of age group, increased. This is consistent with developmental theories regarding children's understanding of emotions (Harter & Whitesell, 1989). Future research, however, is needed to explore the specific reasons *why* children show a tendency to respond at the ex-

tremes when providing ratings of emotional states.

We also hypothesized that there would be age-related differences in children's use of Likert-type rating scales when responding to a subjective, feelings questionnaire. Consistent with the results already described, younger children (i.e., 5- to 6-year-olds) had significantly more extreme scores than the 7- to 9-year-olds and the 10- to 12-year-olds. The size of this effect was large ($\eta^2 = .26$) (Cohen, 1988). In addition, the results showed that the tendency to respond at the extremes was correlated across tasks (with shared variance ranging from 22% to 26%) and was significantly related to children's tendency to respond at the extremes on the feelings questionnaire. Not surprisingly, the highest correlation was between the social objective and subjective tasks, the tasks that were most similar in format (i.e., both involved the reading of brief stories without a visual accompaniment). Of note, although extreme scores were generally quite low for the physical task, children's tendencies to respond at the extremes on the social objective and subjective tasks were significantly related to their tendency to respond at the extremes even on the physical task. Therefore, these results can be interpreted to indicate that younger children do respond in an extreme manner when providing ratings on a self-report questionnaire. Although the "truth" of the children's ratings on this questionnaire is not possible to establish, the fact that their tendency to respond at the extremes on the questionnaire was related to their tendency to respond at the extremes on the investigator-manipulated tasks, where the "truth" was established, lends itself to the inference that the younger children may not have been using the rating scales appropriately in responding to the feelings questionnaire.

Interestingly, there were no differences in children's extreme scores as a function of the number of options included in the response array. Children who used the three-choice options had similar extreme scores on all three tasks and the feelings questionnaire as did children who used the five-choice options. The corresponding effect sizes for this finding were small ($\eta^2 = .03$ and $.05$); therefore, lack of power is an unlikely explanation for the failure to find differences. Simplifying the number of response options available to children does not appear to reduce their tendency to respond at the extremes of rating scales.

To ensure that the pattern of results could be attributed to the tendency of younger children to

respond at the extremes of rating scales, and not simply the hesitancy of older children to use the extremes, we analyzed children's responses on the filler items for each task, which were manipulated so that the correct answer would be, in fact, at the extremes. Results from these analyses showed that there were no differences in children's extreme scores as a result of age or number of response choices for the three tasks. This supports the view that it is the younger children's inability to use the middle points of rating scales that best accounts for the age differences.

The results of this research have potentially substantial implications for the interpretation of self-report ratings from children. For example, research examining age differences in children's self-reports of pain have found that younger children generally report more pain than older children (e.g., Goodenough et al., 1997). These findings have been interpreted as representing a true developmental difference in children's experiences of pain. However, the finding of significantly higher pain ratings among younger children might simply be attributable to their tendency to respond at the extremes of rating scales (which might elevate their ratings compared to older children). Similarly, research examining agreement between parent and children reports of pain has shown low levels of agreement between parent and child ratings, with parents underestimating their children's pain intensity (e.g., Chambers, Reid, Craig, McGrath, & Finley, 1998). These results have been interpreted as indicating that parents are generally not very accurate judges of their children's pain. However, if younger children are indeed responding at the extremes of rating scales (which might then result in higher child ratings relative to parent ratings), parents' ratings might be in fact the more accurate source of information in comparison to their children's ratings. Clearly, the results of this study confirm that younger children do respond in an extreme manner when using rating scales to report on emotional states. The degree to which previous research findings have been influenced by this tendency is an issue that deserves future research attention.

There were several limitations to this research. First, children's ages were used as an approximation of their developmental level and cognitive ability. Future research should incorporate more fine-grained descriptions of children's performance on Piagetian tasks and verbal and nonverbal intellectual abilities and explore how these measures relate

to children's tendencies to respond at the extremes on rating scales. Second, the children used in this study were medically and psychologically healthy children from middle to upper class families. However, we speculate that, if the effects discussed were present in these healthy, well-adjusted children, children who are actually experiencing medical or psychological distress, as is often the case in pediatric psychology research, might show an even greater tendency to respond at the extremes of rating scales. In addition, the physical task consisted of both verbal and visual information, whereas the other tasks consisted of only verbal information. Children's good performance on the physical task might be attributed to the paired sources of information provided for the children to base their rating on, as opposed to the other tasks being more specific to emotional content.

In general, further research is needed to examine factors that might minimize the tendency of children to respond at the extremes of rating scales and maximize the likelihood that children provide accurate and valid self-reports on questionnaires. In keeping with the majority of research which relies on children's self-report ratings, other than to ensure familiarity with their response options, this study did not employ an opportunity for children to practice or train to use the rating scales. Research is needed to examine whether a brief training procedure could eliminate younger children's tendencies to respond at the extremes. In addition, research should explore different methods of eliciting children's responses to questionnaire items. The availability of appropriate visual aids (e.g., representing the response choices with different sized bars or circles) might enhance children's abilities to use rating scales. Some investigators have explored the use of faces varying in degrees of distress and colors varying in intensity in the place of word descriptors in ratings scales (e.g., McGrath, de Veber, & Hearn, 1985), although no research to date has clearly established that this method is superior to the more traditional verbal response choices. Further, Harter (1982), in her rating scale measuring

self-concept, asks children questions with two logically opposed descriptions (e.g., "Some kids often forget what they learn" but "Other kids remember things easily"), and children must first select one or the other and then judge it to be "really true of me" or "sort of true for me." Although this questionnaire has excellent psychometric properties and is among the most highly regarded measures of children's self-esteem, the specific measurement properties of this response format have not been explicitly tested. However, it seems likely that, by providing only two response options at a time, children might be able to provide more accurate ratings.

In sum, the results of this study indicate that young school-age children respond in an extreme manner when rating emotional, psychological states and that this tendency might have an erroneous and invalid impact on the interpretation of children's self-reports. Simplifying the number of response options available to children apparently does not reduce this tendency. Although future research is needed, researchers and clinicians should take this tendency into account when analyzing and interpreting the self-report responses of young children.

Acknowledgments

This research was supported by a Medical Research Council of Canada (MRC) Doctoral Award awarded to C. T. Chambers and a National Health and Research Development Program (NHRDP) grant awarded to C. Johnston. We thank members of the UBC Parenting Research Lab for their assistance with subject recruitment, Robert Chambers for drawing the cartoon pictures used in the physical task, Annikka Kauppinen for her assistance with data collection, and Dr. Julian Somers for allowing us to use the UBC Psychology Clinic for testing.

Received June 14, 1999; revisions received March 25, 2000, and September 4, 2000; accepted September 4, 2000

References

- Banerjee, M. (1997). Peeling the onion: A multilayered view of children's emotional development. In S. Hala (Ed.), *The development of social cognition* (pp. 241-272). East Sussex, UK: Psychology Press Ltd.
- Chambers, C. T., & Craig, K. D. (1998). An intrusive impact of anchors in children's faces pain scales. *Pain*, 78, 27-37.
- Chambers, C. T., Reid, G. J., Craig, K. D., McGrath, P. J., &

- Finley, G. A. (1998). Agreement between child and parent reports of pain. *Clinical Journal of Pain, 14*, 336–342.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Donaldson, S. K., & Westerman, M. A. (1986). Development of children's understanding of ambivalence and causal theories of emotion. *Developmental Psychology, 22*, 655–662.
- Gelman, R., & Baillargeon, R. (1983). A review of some Piagetian concepts. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology, volume 3: Cognitive development*. New York: Wiley Press.
- Goodenough, B., Kampel, L., Champion, G. D., Laubreaux, L., Nicholas, M. K., Ziegler, J. B., & McNerney, M. (1997). An investigation of the placebo effect and age-related factors in the report of needle pain from venepuncture in children. *Pain, 72*, 383–391.
- Harris, P. L. (1983). Children's understanding of the link between situation and emotion. *Journal of Experimental Child Psychology, 36*, 490–509.
- Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development, 53*, 87–97.
- Harter, S., & Buddin, B. J. (1987). Children's understanding of the simultaneity of two emotions: A five-stage developmental acquisition sequence. *Developmental Psychology, 23*, 388–399.
- Harter, S., & Whitesell, N. R. (1989). Developmental changes in children's understanding of single, multiple, and blended emotion concepts. In C. Saarni & P. L. Harris (Eds.), *Children's understanding of emotion* (pp. 81–116). Cambridge: Cambridge University Press.
- La Greca, A. M. (1990). *Through the eyes of the child: Obtaining self-reports from children and adolescents*. Boston: Allyn and Bacon.
- Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology, 22*, 37–49.
- McGrath, P. A., de Veber, L. L., & Hearn, M. J. (1985). Multidimensional pain assessment in children. In H. L. Fields, R. Dubner, & F. Cervero (Eds.) *Proceeding of the Fourth World Congress on Pain. Advances in pain research and therapy, vol. 9*. New York: Raven Press.
- Miller, D. C. (1977). *Handbook of research design and social measurement*. New York: MacKay.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press.
- Rodrigue, J. R., Geffken, G. R., & Streisand, R. M. (2000). *Child health assessment: A handbook of measurement techniques*. Boston: Allyn and Bacon.
- Surber, C. F. (1984). Issues in using quantitative rating scales in developmental research. *Psychological Bulletin, 95*, 226–246.
- von Baeyer, C., Carlson, G., & Webb, L. (1997). Underprediction of pain in children undergoing ear piercing. *Behavioural Research and Therapy, 35*, 399–404.