

Direction of wording and responses to items in oral health-related quality of life questionnaires for children and their parents

Locker D, Jokovic A, Allison P. Direction of wording and responses to items in an oral health-related quality of life questionnaires for children and their parents. Community Dent Oral Epidemiol 2007; 35: 255–262. © 2007 The Authors. Journal compilation © 2007 Blackwell Munksgaard

Abstract - Objectives: In order to minimize acquiescence response set bias, it is often recommended that questionnaires measuring attitudes, behaviours or health states contain items worded positively and negatively. It has also been suggested that when measuring health status this approach means that both the negative and positive dimensions of health can be assessed. This study aimed at assessing the performance of negatively and positively worded items in questionnaires to measure child and parent perceptions of child oral healthrelated quality of life. Methods: Both the child and parent questionnaire included four pairs of items, one negatively worded and one positively worded, that assessed eating, appearance, oral self-care and self-confidence. The response format was a five-point Likert frequency scale with a 'Don't know' option. Prior to analysis, the positive items were reverse coded. The relative performance of the two sets of items was assessed by means of comparisons of the proportions with 'Don't know' responses or missing values, mean item scores and proportions with the two highest frequency codes. Kappa statistics and intraclass correlation coefficients were used to assess the agreement between the negative and reverse-coded positive items and scores and the agreement between child and parent pairs. Factor analysis was used to determine if the two sets of items were measuring the same underlying construct. Results: The study was completed by 91 Canadian children and 100 parents (91 child-parent pairs) recruited from clinics treating paediatric, orthodontic and oro-facial conditions. The positively worded items elicited substantially more 'Don't know' responses or missing values than the negatively worded items and failed to discriminate between groups. In addition, mean item scores and proportions with the highest frequency codes were substantially larger for reverse-coded positive than negative items. Agreement between pairs of items was slight. Child-parent agreement was substantial for a scale constructed from the negative items but only moderate for the positive items. Factor analysis revealed that the two sets of items loaded onto different factors. Conclusions: The performance of the positively worded items was unsatisfactory and their use in oral health-related quality of life indexes, either to reduce response set or assess positive oral health, is at best questionable.

David Locker¹, Aleksandra Jokovic¹ and Paul Allison²

¹Community Dental Health Services Research Unit, Faculty of Dentistry, University of Toronto, Toronto, ²Faculty of Dentistry, McGill University, Montreal, Canada

Key words: acquiescence response set; children; oral health-related quality of life; positive health; response bias

Dr David Locker, Faculty of Dentistry, University of Toronto, 124 Edward Street, Toronto, Ontario, Canada M5G 1G6 Tel: +1 416 979 4907 ext 4490 Fax: +1 416 979 4936 e-mail: david.locker@utoronto.ca

Submitted 11 August 2005; accepted 30 November 2005

Locker et al.

Many texts describing the construction and design of psychological and behavioural questionnaires discuss biases that may arise because of various response effects. The two most often described are acquiescence response set, sometimes called affirmation or agreement bias, and social desirability bias (1–4).

Acquiescence response set is a tendency to give an identical response to each item irrespective of its content that is induced by a sequence of questions in which the direction of wording and response formats is the same. The answers reflect respondents' general feelings about the topic of investigation, rather than the specific issue addressed by each question (5). Social desirability bias is an unconscious tendency for a respondent to give a socially acceptable response to items that contain a degree of threat because they violate social norms, such as questions about drinking, smoking, sexual behaviour or even dental visiting. This tendency is often considered to be a personality trait and differs from a situation in which a respondent is deliberately attempting to mislead by 'faking good' (2).

It has been suggested that both of these response effects can compromise the validity of a scale and result in the under- or overestimation of the relationships between dependent and independent variables (6). Consequently, many experts suggest that questionnaires need to be designed in ways that eliminate or at least minimize these types of bias. One way of dealing with acquiescence response effects is to vary the wording of the questions so that some are phrased negatively and some positively (2, 3). By altering the direction of the wording, the assumption is that respondents will consider each question more carefully rather than giving a more generalized response (3).

Clearly, many of the items comprising oral health-related quality of life (OHQOL) questionnaires can be phrased negatively or positively. However, the GOHAI is the only OHQOL measure developed to date that uses both negative and positive items in an attempt to minimize acquiescence response set (7), with nine negatively worded and three positively worded items. McGrath et al. (8) have suggested that such scales are preferable to ones containing only negative items as they can assess both the negative and positive dimensions of oral health. Consequently, they provide information on the health status of those free of disease and disability (9) and provide a more comprehensive assessment of the way in which oral disorders both compromise and enhance the quality of life.

While the use of both positively and negatively worded items is common in questionnaires designed to assess attitudes, psychological states and behaviours, the practice has been criticized on a number of grounds (5). DeVellis (1) has suggested that there may be a price to pay if the direction of items is changed throughout a questionnaire. A respondent may become confused by such changes and this may lead to bias or errors in responding. Secondly, responses to questionnaire items are often influenced by question wording and formats (10, 11), so that positively and negatively worded versions of the same question can give rise to very different estimates of the prevalence of the event being measured. Cohen et al. (12) and Foddy (4) provide examples of how changes in the direction of wording can have a major effect on the distribution of responses and survey estimates.

A third problem is that the use of both positively and negatively worded items means that the coding of one or the other must be reversed in order to calculate overall scale scores. The use of reverse coding rests on two assumptions: (a) that positively and negatively worded items measure the same underlying construct (13), and (b) that the categories of Likert-type response formats are symmetrical and equi-distant (3, 14).

These assumptions were examined in a study that used the Zung Self-rated Depression Scale (15) and Speilberger's State Trait Anxiety Index (16). Both measures are balanced with equal numbers of symptom-positive and symptom-negative items (13). The consequences of using both negative and positive statements was examined by comparing the mean scores of positively worded items with those of the negative items after reversal of the latter's response codes. It was assumed that if the two types of item were assessing the same underlying construct, then the mean scores derived from the positive and negative items would be similar. However, for both measures the reversed-negative items resulted in significantly higher mean scores than the positive items. Furthermore, it was assumed that if the same construct was being measured irrespective of the direction of the wording, then factor analysis would identify factors that were independent of the way in which questions were phrased. This was not the case; the positive and negative items loaded onto different factors. Similar results were obtained from a factor analysis of the General Health Questionnaire (17), a measure of symptoms potentially indicative of psychiatric disorder that also has negatively and positively worded symptom statements.

Studies have not been undertaken to examine the effect of direction of wording on responses to items in OHQOL scales and indexes, given the lack of balanced measures or measures in which negative and positively worded versions of the same item have been used. Consequently, this paper explores this issue in the context of a study of child oral health-related quality of life (COHQOL).

Methods

The data on which this paper is based were collected as part of a study to assess cross-cultural variations in the OHQOL of children aged 10–14 years. The study employed the conceptual framework, methodology and preliminary item pool used by Jokovic et al. (18, 19) in developing the COHQOL questionnaires, and produced modified forms of the child and parental components of the COHQOL questionnaire. These modified forms were renamed as the Child Oral Health Impact Profile (COHIP) (20).

In order to assess the effects of the direction of question wording on questionnaire responses and the feasibility of measuring both negative and positive oral health, four items dealing with eating, oral self-care, appearance and self-confidence were written in both negative and positively worded forms. These asked about the frequency of negative and positive events and experiences relative to these domains over the previous three months. The response format of all questions was a Likert-type frequency scale with the following options: never = 1; almost never = 2; sometimes = 3; fairly often = 4; almost all of the time = 5. A 'Don't know' response was also allowed for each item. The wording of the negative and positive

versions of the questions used in the parent questionnaire is shown in Table 1 along with the order in which the questions appeared on the questionnaire. The wording of the questions in the child questionnaire was identical except that the words 'your child' and 'his/her' were changed to 'your' or 'you', as appropriate.

Study subjects were children and their parents attending clinical settings in Montreal, Canada for the treatment of various oral and oro-facial disorders. As all participants were French-speaking, the questionnaires were translated into French using the comprehensive forward-backward technique recommended by Behling and Law (21). Child and parent participants completed questionnaires independently and were asked not to discuss their answers. The subjects were classified into one of three clinical groups; paediatric dental patients, orthodontic patients and patients with oro-facial conditions, predominantly cleft lip and palate with a small number of non-syndromal craniofacial anomalies. All children included in the study were otherwise healthy. As both children and their parents were included, the data provided the opportunity to compare the performance of negatively and positively worded items for patients and proxy informants (22) and to assess effects on agreement between children and their parents.

All aspects of the study including methods of obtaining informed consent and assent from parents and children were judged to be ethical by McGill University's Research Ethics Committee and the University of Toronto Ethics Review Office.

Data analysis

Separate analyses were conducted for child and parent (proxy) participants. The aim of the analyses was to determine (i) if estimates derived from the negative and reverse-positive pairs were similar or

During the last 3 months, how often has your child...because of his/her teeth, mouth or face? Negative Question Question Question item number Positive item number Had difficulty eating foods he/ Been able to eat foods he/ 15 29 Eating she would like to eat she wants to eat Appearance Felt that he/she was not good looking 22 Felt that he/she was attractive 31 (good looking) Oral self-care^a Had difficulty keeping his/her teeth clean 32 Been able to keep his/her teeth clean 17 Self-confidence Felt shy or withdrawn Been confident 14 18

Table 1. Negatively worded and positively worded items from the parental questionnaire

Response format: 1, Never; 2, Almost never; 3, Sometimes; 4, Fairly often; 5, Almost all of the time; 8, Don't know. ^aSuffix not used with these items.

different, (ii) the extent of agreement between pairs of items, and (iii) whether or not the positive and negative versions of the items appeared to be measuring the same underlying construct. Prior to analysis, the four positive items were reverse coded, i.e. 1 changed to 5, 2 to 4, 3 to 3, 4 to 2 and 5 to 1.

First, the percent of subjects with 'Don't know' or missing responses for each item were calculated. The mean score for each item was calculated and scores for the negative and reversed-positive items compared using paired *t*-tests. The proportion of subjects with response codes 4 or 5 were also calculated for each item and the pairs compared using McNemar's test. This cut-off point was chosen as it is the one usually used when calculating simple count scores for OHQOL scales. Agreement between the dichotomized negative and reverse-positive item pairs was assessed using the kappa statistic.

The response codes of the four negatively worded items were then summed, as were the response codes of the four reverse-coded positively worded items, to create two short scales. Mean scale scores, the per cent with one or more 'Don't know' responses or missing values were compared and agreement between the scores was assessed using the intraclass correlation coefficient (ICC).

For each of the eight items, the agreement between child and parental responses was examined using the kappa statistic and the agreement between the negative and positive scales examined using the ICC. Both the weighted kappa and ICC were interpreted using the guidelines of Landis and Koch (23). These were: <0.20 – poor; 0.21–0.40 – fair; 0.41–0.60 – moderate; 0.61–0.80 – good; and 0.81–1.0 – very good to perfect. Negative values indicate agreement worse than chance (24).

Finally, the eight items for children and the eight for parents were subjected to factor analysis to determine if the factors produced were independent of the direction of wording.

Results

Participants

Data were collected from 91 children and 100 parents (91 child–parent pairs). Of the children, 50 were girls and 41 boys and their mean age was 11.6 years. Fifty-six were in the paediatric dental group, 15 in the orthodontic group and 20 in the oro-facial group. The majority of the parental

participants (72.4%) were female. The mean number of years of education of the parents was 13.5 years.

'Don't know' responses and missing values

Overall, 16.3% of children had one or more 'Don't know' responses or missing values for the negatively worded items while 39.1% had one or more 'Don't know' or missing values for the reversed positively worded items. For parents, the percentages were 10.2% and 49.0%, respectively. Table 2 shows these percentages for the individual items. Significant differences were found for one of the four item pairs for children and two for parents. Children and parents were similar with respect to the proportions giving a 'Don't know' or missing response when negatively worded items were used. When positively worded items were used to elicit parental reports of psychological states, such as feelings concerning appearance and selfconfidence, they were particularly prone to 'Don't know' responses or missing values. Almost onethird responded in this way to the positively worded item about self-confidence.

Mean item scores and proportions with response codes 4 or 5

For children, the mean scores for the reversedpositive items were significantly higher than the negative items for three of the four item pairs (eating, appearance and self-confidence) (Table 3). The proportions with response codes of 4 or 5 were also significantly higher for the reversed-positive items in these three domains. For parents, the mean scores of the reversed-positive items were significantly higher than the mean scores of the negative

Table 2. Percentage of respondents with 'Don't know' responses or missing values

	0	
Question	Negatively worded item	Reverse-coded positively worded item
Children ($n = 91$) (%)	
Eating	5.5	13.2
Appearance	6.6	13.2
Oral self-care	5.5	4.4
Self-confidence	7.7*	20.9
Parents ($n = 100$) (%)	
Eating	3.1	4.1
Appearance	8.2**	29.6
Oral self-care	3.1	1.0
Self-confidence	6.1**	31.7

Differences in proportions between negative and positively worded item pairs statistically significant: *P < 0.05; **P < 0.001; McNemar's test.

Table 3. Mean item scores and	d percent with res	sponse codes 4 or 5 for	negative items and	reverse-coded	positive items
			()		

Question	Negatively worded item	Reverse-coded positively worded item	
Children			
Eating $(n = 74)$	1.72 (SD 1.09)* 8.1% [†]	2.43 (SD 1.68) 29.7%	
Appearance $(n = 74)$	1.93 (SD 1.20)** 9.5% [†]	3.43 (SD 1.60) 55.7%	
Oral self-care $(n = 83)$	1.76 (SD 1.01) 4.8%	1.73 (SD 0.96) 9.6%	
Self-confidence ($n = 65$)	1.51 (SD 0.92)** 4.6% [†]	2.94 (SD 1.67) 44.6%	
Parents			
Eating $(n = 93)$	1.65 (SD 1.00)** 4.3% ^{††}	2.52 (SD 1.73) 34.4%	
Appearance $(n = 69)$	1.62 (SD 1.05)** 4.3% ^{††}	2.94 (SD 1.50) 30.4%	
Oral self-care $(n = 96)$	2.19 (SD 1.14)* 9.5%	1.81 (SD 1.27) 8.3%	
Self-confidence ($n = 69$)	1.67 (SD 1.05)** 5.8% ^{††}	2.55 (SD 1.65) 30.4%	

Differences in mean values: *P < 0.01, **P < 0.001; paired *t*-test. Differences in proportions: [†]P < 0.01, ^{††}P < 0.001; McNemar's test.

items for all of the item pairs. Again, the proportions with response codes of 4 or 5 were also significantly higher for the reversed-positive items for the eating, appearance and self-confidence questions. For both children and parents, the differences in the proportions with the highest response codes was substantial, varying from a threefold to almost a 10-fold difference.

These differences at the level of individual items are reflected in mean scale scores. The reversedpositive scale mean for children was 11.5 (SD = 3.6), significantly higher than the negative scale mean of 7.0 (SD = 2.7) (P < 0.001; paired t-test). The mean values for parents were 9.1 (SD = 3.9) and 7.1 (SD = 3.1) respectively (*P* < 0.01; paired *t*-test).

Agreement between item pairs and scales

Kappa statistics for agreement between the negatively worded and reversed positively worded pairs of items were 0.11 and under for children (Table 4). Two item pairs showed poor agreement and two were close to zero and showing no agreement. For

Table 4. Kappa statistics for agreement between negatively worded and reversed positively worded pairs of items

Question	Kappa (SE)
Children	
Eating $(n = 74)$	0.10 (0.10)
Appearance $(n = 74)$	0.09 (0.06)
Oral self-care $(n = 74)$	0.11 (0.15)
Self-confidence ($n = 65$)	-0.23 (0.06)
Parents	
Eating $(n = 93)$	-0.023 (0.05)
Appearance $(n = 69)$	0.19 (0.01)
Oral self-care ($n = 96$)	0.29** (0.16)
Self-confidence ($n = 69$)	0.16* (0.01)

*P < 0.05, **P < 0.01.

parents, one pair of items showed fair agreement (K = 0.29) (Table 4) with the remaining three showing poor agreement.

The ICC for agreement between the negative item scale scores and the reversed-positive item scale scores was -0.73 (95% CI -2.0 to 0.10) for children, indicating agreement worse than chance, and 0.47 (95% CI 0.07 to 0.70) for parents, indicating moderate agreement.

Child-parent agreement

Table 5 shows the kappa statistics for agreement between children and parents for each of the eight items. These ranged from -0.05 (NS) to 0.37 (P < 0.001) with the negatively worded items showing better agreement than the reverse-coded positively worded. Agreement was poor except for one item, the negatively worded item on oral selfcare where it was moderate.

The ICC for child-parent agreement for the negatively worded scale scores was 0.70 (95% CI 0.51 to 0.81) indicating substantial agreement, and

Table 5. Kappa statistics for agreement between children and parents for each of the eight items

Question/wording	Kappa (SE)
Eating	
Negatively worded $(n = 83)$	0.18 (0.19)
Reversed positively worded ($n = 75$)	0.02 (0.12)
Appearance	
Negatively worded $(n = 79)$	0.16 (0.17)
Reversed positively worded ($n = 57$)	0.09 (0.12)
Oral self care	
Negatively worded $(n = 84)$	0.37 (0.20)*
Reversed positively worded ($n = 86$)	0.07 (0.13)
Self-confidence	
Negatively worded $(n = 79)$	-0.05 (0.02)
Reversed positively worded ($n = 53$)	0.01 (0.13)

*P < 0.001.

Locker et al.

was 0.41 (95% CI 0.21 to 0.72) for the reversed positively worded scale scores, indicative of only fair agreement (22).

Factor analysis

Factor analysis of the child data produced a threefactor solution that accounted for 57% of the variance. Three negatively worded items (eating, appearance and self-confidence) loaded on the first factor, three positively worded items (self-confidence, appearance and oral self-care) loaded on the second and one negative item (oral self-care) and one positive item (eating) loaded on the third. The analysis of the parental data resulted in a threefactor solution, accounting for 62%% of the variance. All four negative items loaded on the first factor, three positive items loaded on the second and one positive item on the third. These results suggest that the negatively worded and positively worded items are not measuring the same underlying construct.

Discussion

This paper has reported the results of a preliminary investigation of the use of negatively and positively worded items in OHQOL questionnaires for children and their parents. We studied pairs of items in which one of each pair assessed the frequency of a negative event (negatively worded), while the other assessed the frequency of the event phrased in positive terms (positively worded). As the majority of OHQOL questionnaires consist exclusively of negatively worded items, we have used these as the 'gold standard' against which their positively worded counterparts may be judged. As the study was confined to four pairs of items assessing a limited number of oral health domains and to participants recruited from clinical settings, the study represents a first step in what should be a broader enquiry into the effect of question wording on estimates derived from OHQOL questionnaires.

One clear outcome of this study is that the positively worded items elicited substantially more 'Don't know' responses or missing values than the negatively worded items. Two-fifths of the children and one-half of the parents provided these responses for one or more of the former. Such responses are problematic. Respondents with these responses must be excluded from the analysis or these values must be imputed in some way (25). Moreover, the high prevalence of 'Don't know' responses suggests that both groups of participants had trouble either understanding or providing responses to these questions and that parents as proxies found these particularly difficult. This may be because positively worded items address events that are either more subtle or less obvious than negatively worded items (2) or that they are at odds with conventional ways of thinking about and framing events relating to health. That is, we may notice the days when we have pain but may not notice the days on which pain is absent. As Dingwall (26; p. 62) has commented with respect to health: 'Given the unobtrusiveness of what is normal, by virtue of its very normality, the unusual may be easier to identify'. It may also be the case that, as a consequence, the cognitive processes involved in responding to the type of positive items assessed here is more complex than in responding to their negative counterparts.

A second indicator of the problematic performance of the positively worded items and the scale from which they were constructed was that they resulted in higher mean values and prevalence rates than their negatively worded counterparts. Consequently, items that are phrased positively may give an inflated estimate of the extent to which physical, psychological and social functioning is compromised by oral disorders. They may then be less useful than items documenting negative events in descriptive studies of the OHQOL of various populations.

However, perhaps the observation of most concern is that the level of agreement between negatively worded items and reverse-coded positive items is very low and that agreement between scores derived from scales constructed from these items is also low. This was particularly the case for children where the ICC for agreement between the two scales approached -1.0. Agreement was better for parents but was at best moderate. It is, of course, possible that poor agreement may be due to differences in the wording of negative and positive items, so that one is not the exact opposite of the other, or the fact that the response scale is unbalanced and the categories it includes not equidistant. For example, one end of the scale is 'Never' which is an absolute, while the other end is 'Almost all the time' which is not. However, it was also the case that agreement between child and parent pairs was also better with the negative as opposed to the reverse-coded positive scale scores so that poor questionnaire technique may not explain our observations.

These results suggest that the negative and positive items may not be measuring the same underlying construct even though they were written with this intention in mind. This conclusion is supported by the factor analyses where the negative and positively worded items loaded onto separate factors. This indicates that the positively worded items are more highly correlated with each other than they are with their negatively worded counterparts. This raises the intriguing question of what exactly is being measured when people respond to items concerning oral health-related negative and positive events.

These findings tend to support a body of literature that has questioned the use of bi-directional items (5). This research has raised questions about the internal reliability, descriptive statistics and factor structures of balanced scales and the ability of respondents to cope with changes in the direction of item wording. However, this research has examined scales that consist of directly worded and what might be called 'true' negatively worded items. Consequently, what we mean by the terms 'negatively worded' and 'positively worded' as used in this paper needs clarification.

For the sake of brevity, we have characterized items such as 'How often in the past three months have you been unhappy' as negatively worded items. However, items such as this are in fact a positive item about a negative event. That is, they assess the presence of an adverse experience. The reverse item, 'How often in the past three months have you been happy' is a positively worded item about a positive event. Again, it assesses the presence of a favourable experience. The negatively worded versions of these items would be; 'How often in the past three months have you not been unhappy' and 'How often in the past three months have you not been happy'. These assess the absence of negative or positive experiences. There appears to be a consensus that questionnaire items containing the word 'not' are best avoided. The literature on acquiescence response effects and the use of items worded in opposite directions does not always make this distinction clear.

Nevertheless, the limited data presented here suggest that there might be problems when using items that we have characterized as positively worded, and these problems are akin to those encountered when using balanced scales. Barnette (5) has advanced the view that acquiescence response set is uncommon and that for most surveys it is not necessary to guard against this source of bias, particularly when participation is voluntary and respondents can be expected to respond to the best of their ability. He suggests that unless there is some compelling reason to use a balanced scale then a measure in which all items are unidirectional is to be preferred. Bradburn (6) has suggested that even if such acquiescence response effects are observed, they should be regarded as true differences between individuals rather than biases that need to be eliminated.

If an investigator suspects that response effects may be a problem there are other solutions, such as reversing the response coding, rather than reversing the direction of some items, that improve rather than diminish the performance of a questionnaire (5). Our data suggest that, in the context of health-related quality of life questionnaires, what we have called positively worded items do not function particularly well. If included in order to reduce acquiescence, then a solution might be to discard the data they provide and construct scores from those items that focus on more readily recalled and less ambiguous negative events. These conclusions are preliminary and further research including qualitative studies is necessary to investigate respondents' understandings of items in OHQOL questionnaires, the meaning of responses to positively and negatively worded items and the ways in which wording may affect their responses to those items.

Acknowledgments

This work was supported by NIDCR grants: R01DE13732 and R21DE13721. We would like to thank the following for their contribution to the international collaborative study which developed the questionnaire used in the study: Anna Leao, Colman McGrath, Murray Thomson, John Broughton, Tim Newton, Berta Prahl, Martine Hennequin, Hillary Broder, Susan Reisine, David Gibbons and Sue Naidoo.

References

- 1. De Vellis R. Scale development: theory and applications. Newbury Park, CA: Sage Publications; 1991.
- 2. Streiner D, Norman G. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press; 1989.
- 3. Bowling A. Research methods in health: investigating health and health services. Buckingham, UK: Open University Press; 1997.

Locker et al.

- 4. Foddy W. Constructing questions for interviews and questionnaires: theory and practice in social research. Cambridge: Cambridge University Press; 1993.
- 5. Barnette J. Effect of stem and Likert response option reversals on survey internal consistency: if you feel the need there is a better alternative to using those negatively worded items. Educ Psychol Measurement 2000;60:361–70.
- 6. Bradburn N. Response effects. In: Rossi P, Wright J, Anderson A, editors. Handbook of survey research. Orlando, FL: Academic Press Inc; 1983.
- Atchison K, Dolan T. Development of the Geriatric Oral Health Assessment Index. J Dent Educ 1990;54:680–7.
- McGrath C, Broder H, Wilson-Genderson M. Assessing the impact of oral health on the quality of life of children: implications for research and practice. Community Dent Oral Epidemiol 2004: 32:81–6.
- 9. Bowling A. Measuring health: a review of quality of life measurement scales. Buckingham: Open University Press; 1991.
- Benson J, Hocevar D. The impact of item phrasing on the validity of attitude scales for schoolchildren. J Educ Meas 1985;22:231–40.
- 11. Schwartz N. Self-reports: how questions shape the answers. Am Psychol 1999;54:93–105.
- Cohen G, Forbes J, Garraway M. Can different patient satisfaction methods yield consistent results? Br Med J 1966;313:841–844.
- Schotte C, Maes M, Cludyts ?, Cosyns P. Effects of affective semantic mode of item presentation in balanced self-report scales: validity of the Zung selfrating depression scale. Psychol Med 1996;26:1161–8.
- 14. Keller S, Ware J, Gandek B, Aaronson N, Alsonso J et al. Testing the equivalence of widely used response choice labels: results from the IQOLA project. J Clin Epidemiol 1998;31:933–44.
- 15. Zung W. A self-rating depression scale. Arch Gen Psychiatry 1965;12:63–70.

- 16. Speilberger C, Edwards C, Mantoun J, Lushene R. The State-Trait Inventory. Windsor: NFER-Nelson; 1987.
- 17. Huppert F, Whittington J. Evidence for the independence of positive and negative well-being: implications for quality of life assessment. Br J Health Psychol 2003;8:107–22.
- Jokovic A, Locker D, Stephens M, Kenny D, Tompson B. Validity and reliability of a measure of child oral health-related quality of life. J Dent Res 2002, 81:459–63.
- 19. Jokovic A, Locker D, Stephens M, Kenny D, Tompson B, Guyatt G. Measuring parental perceptions of child oral health-related quality of life. J Public Health Dent 2003;63:67–72.
- 20. Broder H, Janal M, Wilson-Genderson M, Reisine S, Phillips C. Reliability and validity of the Child Oral Health Impact Profile. J Dent Res 2005;84 (Special Issue A): 2562.
- 21. Behling O, Law K. Translating questionnaires and other research instruments: problems and solutions series. Thousand Oaks, CA: Sage Publications; 2000.
- 22. Jokovic A, Locker D, Stephens M, Guyatt G. Agreement between mothers and children in rating child oral health-related quality of life. Community Dent Oral Epidemiol 2003;31; 335–43.
- 23. Landis J, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
- 24. Altman G. Practical statistics for medical research. Boca Raton, FL: Chapman & Hall/CRC; 1991
- 25. Jokovic A, Locker D, Guyatt G. How well do parents know their children? Implications for proxy reporting of child health-related quality of life. Qual Life Res 2004;13:1297–1307.
- 26. Dingwall R. Aspects of illness. London: Martin Robertson; 1976.

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.