

Downloaded from UvA-DARE, the institutional repository of the University of Amsterdam (UvA)
<http://hdl.handle.net/11245/2.102474>

File ID	uvapub:102474
Filename	Chapter 2: Item-order effects in the Oral Health Impact Profile
Version	unknown

SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type	PhD thesis
Title	Some fundamental issues in oral health-related quality of life research
Author(s)	J.M. Kieffer
Faculty	ACTA
Year	2012

FULL BIBLIOGRAPHIC DETAILS:

<http://hdl.handle.net/11245/1.372260>

Copyright

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content licence (like Creative Commons).

CHAPTER 2
Item-Order Effects in the Oral Health Impact Profile

INTRODUCTION

In social research, self-report questionnaires are commonly used to assess people's states, feelings or attitudes. In so doing, it is assumed that the underlying construct to be measured is adequately reflected by the responses given. But it has long been recognized that response format, response alternatives, and item order can influence responses (Weinberger et al. 2006). These so-called 'context effects' can go unnoticed, and when systematic they can have a dramatic effect on the factor structure and on research outcomes. Therefore, it is essential to recognize these potentially harmful effects (Weinberger et al. 2006). Item-order effects fall under context effects because they can occur as a result of the context or conceptual framework surrounding a construct that is intended to be measured by a questionnaire. When people are asked to form a judgment, they first have to define a standard of comparison to measure their opinion against. This standard can be formed from existing information in the memory, but also from stimuli given, for instance, in an earlier item (Sudman et al. 1996). Therefore, the idea of the context of comparison typifies the importance of the relationship between items in a questionnaire. Two types of relationships can be distinguished: part-whole and part-part. In part-whole relationships, the content of one item is often more general than the other, thereby covering part or all of the content of the other item (Schuman et al. 1981). When asked, for instance, if one is generally happy, the answer should be the sum of the answers to questions about specific aspects of happiness (marriage, job, friendships). But when one is not happy in one specific area this does not rule out general happiness *per se* (Schuman et al. 1981). A specific question may, in fact, influence the response to a general question considerably. In part-part relationships between items, both items have the same level of specificity. There is not one more general than the other as in part-whole relationships. Nevertheless, they may have an influence on each other (Schuman et al. 1981). An example is questioning people about the integrity of Bill Clinton and Al Gore. Both items have the same level of specificity. However, the effect of asking first about the integrity of Bill Clinton or first about the integrity of Al Gore could be very different on the response to the later item (Moore 2002). Consequential to both relationships, two types of item-order effects can occur: consistency effects and contrast effects. When consistency effects occur, the response to a later item is closer to the response to an earlier item than it would otherwise be. Thus, when, for instance, a response to an item is extremely positive, the response to a later item will be equally positive. On the other hand, contrast effects occur when the response to one item leads to a more opposed response to another item (Schuman et al. 1981).

When measuring quality of life, the general combination between items is a part-whole combination where respondents are asked questions concerning their overall satisfaction with life and their satisfaction with specific subdomains, such as job satisfaction or satisfaction with their marriage (McClendon et al. 1988). In the present study, oral health-related quality of life (OHRQoL) is measured and therefore the focus will be on the part-whole relationship between more specific and more general items. The Oral Health Impact Profile (OHIP), used in this study, is an instrument with seven subscales, by which the following seven domains are distinguished: Functional limitations, Physical pain, Psychological discomfort, Physical disability, Psychological disability, Social disability and Handicap (Slade et al. 1994). These dimensions are based on a conceptual model of oral health (Locker 1988, Slade et al. 1994), which has its foundation

in the ‘Classification of Impairments, Disabilities and Handicaps’ developed by the World Health Organization. The dimensions are hierarchically ordered so that the impacts described by the dimensions are gradually more disruptive to one’s life. For instance, the first dimension, Functional limitations, describes impacts less grave in nature, such as: ‘Have you had difficulty chewing any foods because of problems with your teeth, mouth or dentures?’, this in contrast with the last dimension, Handicap, which describes more severe impacts on daily living, such as: ‘Have you been totally unable to function because of your teeth, mouth or dentures?’ (Slade et al. 1994).

There is good reason to assume that the structure of the OHIP has a part-whole combination because this instrument includes items with different levels of specificity. Therefore item-order effects could be expected. Hence, the aim of this study is to evaluate the effect of changing the order of the items in the OHIP.

MATERIALS & METHODS

Participants

Subjects were psychology freshmen, and filling out questionnaires was a mandatory part of their course. In total, 237 freshmen of whom 71% were women took part. The study design was approved by the Netherlands Institute for Dental Sciences (IOT) and by the Department of Psychology (UvA). Subjects took part voluntarily, were able to stop at any given time, and they were given the appropriate information concerning the aim and general conclusions of this study.

Materials

The OHIP consists of 49 items. Because the subjects were relatively young, three items concerning dentures were excluded, leaving 46 items. These 46 items are dispersed over the seven subscales: Functional limitations (eight items), Physical pain (eight items), Psychological discomfort (five items), Physical disability (eight items), Psychological disability (six items), Social disability (five items) and Handicap (six items). For each item, respondents were asked how often in the previous 12 months they had experienced a certain problem regarding their teeth or mouth. They responded on a Likert type scale, which was coded as follows: 4, very often; 3, fairly often; 2, sometimes; 1, hardly ever; and 0, never. Thus, lower scores indicate good OHRQoL and higher scores indicate poor OHRQoL.

Method

Two versions of the Dutch OHIP (Van der Meulen et al. 2008) were randomly distributed amongst subjects. The first version (n=119, mean age + standard deviation (SD) = 21.2 yr + 5.4); 70% women) was distributed in its original form, as described under ‘Materials’ (thus having a hierarchical scale-order). In the second version (n=118, mean age + SD= 21.2 yr + 5.0; 71% women) the item order was manipulated by positioning the last three subscales (Psychological disability, Social disability and Handicap), as the first three subscales. The order in which the respondents answered the items could be controlled because these items were presented on a computer screen.

Statistical analyses

Internal consistency was determined by calculating Cronbach's alpha for the total scores and subscale scores for both versions of the OHIP. Multivariate Analyses of Variance was used to assess the difference between the scores on the two versions. Additional non-parametric tests were used to assess differences between item scores and subscale scores. Effect sizes were calculated for significant differences. Because an effect size is typically a distribution-based measure, an alternative effect-size measure was used which converts a z-score into an effect-size estimate by dividing the z-score by the squared size of the total sample (Field 2005). This estimate is also interpreted in terms of the ranges of magnitude of change likely to be clinically meaningful, as follows: <0.2 (small difference), through 0.2-0.8 (moderate difference), to > 0.8 (large difference) (Cohen 1977).

To compare the factorial structure of both versions it would have been appropriate to use a two-group confirmatory factor analysis (CFA), but because of the small sample size relative to the amount of parameters, parameter estimation would have been unreliable. As an alternative, differences between subscale intercorrelations of version 1 and 2 were computed. Before being able to compute these differences, subscale intercorrelations were transformed, for both versions separately, into z-scores by means of the Fisher-transformation (Van den Brink et al. 1998) (see Eqn. 1):

$$Z_F = \frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) \quad (1)$$

After transforming the correlations into z-scores, Eqn. 2 was used to detect significant differences between the subscale intercorrelations:

$$Z = (Z_{F,1} - Z_{F,2}) / \left[\frac{1}{n_1-3} + \frac{1}{n_2-3} \right]^{\frac{1}{2}} \quad (2)$$

RESULTS

The internal consistency was calculated for each version of the OHIP (Table 1). As shown, the internal consistencies remained virtually the same in both versions, and were relatively high considering that a Cronbach's alpha of 0.6 is regarded as acceptable.

Table 1. Internal consistency

	Cronbach's alpha	
	Version 1* (n=119)	Version 2* (n=118)
Total scale [0-184]	0.95	0.98
Functional limitations [0-32]	0.62	0.63
Physical pain [0-32]	0.85	0.88
Psychological discomfort [0-20]	0.76	0.82
Physical disability [0-32]	0.85	0.87
Psychological disability [0-24]	0.86	0.94
Social disability [0-20]	0.87	0.87
Handicap [0-24]	0.87	0.80

Values in square brackets indicate range of possible scores

*Two different versions of the Oral Health Impact Profile (OHIP).

Table 2 presents the mean scores of both versions of the OHIP and mean differences between versions. The mean scores on the total scale and on the subscales were

higher in version 2 than in version 1, and overall they indicate a relatively good quality of life. The means and SDs illustrate substantial skewness and low variances in both versions.

Table 2. Mean total, mean subscale scores, mean differences and ranked mean differences between the two versions of the Oral Health Impact Profile.

	Version 1 (n=119)	Version 2 (n=118)	Mean diff.	Mann-Whitney <i>U</i> -test		
	Mean (SD)	Mean (SD)		z-score ^a	p	95% CI ^b
Total scale [0-184]	12.9 (16.0)	15.7 (22.8)	-2.8	-0.186	0.85	0.844 - 0.858
Functional limitations [0-32]	2.7 (2.7)	2.6 (2.9)	0.1	-0.710	0.48	0.470 - 0.489
Physical pain [0-32]	4.9 (5.1)	5.3 (5.7)	-0.4	-0.354	0.72	0.714 - 0.731
Psychological discomfort [0-20]	1.8 (2.4)	1.9 (3.1)	-0.1	-0.398	0.69	0.677 - 0.695
Physical disability [0-32]	1.2 (3.0)	1.5 (3.8)	-0.3	-0.111	0.91	0.910 - 0.921
Psychological disability [0-24]	1.1 (2.4)	2.5 (4.7)	-1.4	-2.868	<0.01**	0.003 - 0.005
Social disability [0-20]	0.5 (1.7)	0.9 (2.4)	-0.4	-2.055	0.04*	0.034 - 0.042
Handicap [0-24]	0.7 (2.3)	1.0 (2.4)	-0.3	-2.027	0.04*	0.039 - 0.047

Values in square brackets indicate range of possible scores

* Ranked mean difference is significant at $p < 0.05$ (2-tailed).

** Ranked mean difference is significant at $p < 0.01$ (2-tailed).

^a Standardized score of the mean ranked difference

^b Monte Carlo Confidence Intervals (2-tailed).

Multivariate analyses showed a significant difference between the two versions ($F(7) = 2.77$, $p = 0.009$). At subscale level the only significant difference was found on the Psychological disability scale ($F(1) = 8.44$, $p = 0.004$). However, because the assumption of equal variances was violated in several subscales, the difference between the subscales was assessed by means of a non-parametric test, the Mann-Whitney *U*-test. These results are also presented in Table 2 and show the significance level calculated by mean ranked differences. Significant differences were found between the scores on the subscales Psychological disability, Social disability and Handicap. For these three subscales, the scores on version 2 were higher than on version 1. The effect-size difference on the Psychological disability scale was 0.19. On both the Social disability and the Handicap scale the effect size difference was 0.13. These effect sizes indicate small differences.

Analyses of the differences between individual items, using the Mann-Whitney *U*-test, showed that the above mentioned differences on subscale level were caused by one or more items (Table 3). None of the items in the scales Functional limitations, Physical pain, Psychological discomfort and Physical disability show significant score differences between the two versions of the OHIP. However, four out of six items in the Psychological disability scale showed significant differences between scores. Moreover, the Social disability scale and the Handicap scale show significant differences between scores in two out of five and one out of six items, respectively. Effect-size differences between item scores were between 0.13 and 0.23, indicating small to moderate differences.

Table 3. Items with significant differences between the two versions of the Oral Health Impact Profile (OHIP)

	Version 1		Version 2		Mann-Whitney <i>U</i> -test	
	Mean (SD) (n=119)	Mean (SD) (n=118)	Mean diff.	z-score ^a	P	95% CI ^b
Psychological disability						
Has your sleep been interrupted because of problems with your teeth or mouth?	0.2 (0.6)	0.3 (0.8)	-0.2	-2.703	<0.01**	0.005 - 0.008
Have you been upset because of problems with your teeth or mouth?	0.2 (0.5)	0.7 (1.1)	-0.5	-3.574	<0.01**	0.000 - 0.000
Have you had trouble to relax because of problems with your teeth or mouth?	0.2 (0.5)	0.4 (0.9)	-0.2	-2.387	0.02*	0.013 - 0.018
Have you felt a bit embarrassed because of problems with your teeth or mouth?	0.3 (0.6)	0.5 (0.9)	-0.2	-2.204	0.03*	0.024 - 0.031
Social disability						
Have you avoided going out because of problems with your teeth or mouth?	0.1 (0.4)	0.2 (0.6)	-0.1	-2.023	0.04*	0.051 - 0.060
Have you had difficulty doing your daily activities because of problems with your teeth or mouth?	0.1 (0.6)	0.3 (0.9)	-0.1	-2.061	0.04*	0.040 - 0.048
Handicap						
Have you felt your health in general has diminished because of problems with your teeth or mouth?	0.1 (0.3)	0.1 (0.4)	-0.1	-2.092	0.04*	0.048 - 0.057

* Ranked mean difference is significant at $p < 0.05$ (2-tailed).

** Ranked mean difference is significant at $p < 0.01$ (2-tailed).

^a Standardized score of the mean ranked difference

^b Monte Carlo Confidence intervals (2-tailed).

SD, standard deviation

Correlations between the subscales of version 1 and version 2, respectively, were all highly significant, ranging from 0.38 to 0.89, with 20% of the subscale intercorrelations being above 0.7 (Version 1) and ranging from 0.68 to 0.91, with 90% of the subscale intercorrelations above 0.7 (Version 2). Interestingly, version 2 showed systematically higher correlations between subscales than the original version 1. To compare differences between subscale intercorrelations, z-scores were calculated. Differences between the z-scores are presented in Table 4. With the exception of the subscale pairs Functional limitations/ Physical disability, Psychological disability/ Handicap and Social disability/ Handicap, significant subscale intercorrelational differences between both versions were found on all other subscale pairs.

Table 4. Standardized correlational difference between version 1 and 2 of the Oral Health Impact Profile

	z-score difference	1	2	3	4	5	6	7
1	Functional limitations	0						
2	Physical pain	-3.154**	0					
3	Psychological discomfort	-2.295*	-3.876**	0				
4	Physical disability	-0.882	-3.610**	-3.055	0			
5	Psychological disability	-2.819**	-4.339**	-4.514**	-4.271**	0		
6	Social disability	-2.576**	-3.625**	-3.207**	-3.572**	-2.827**	0	
7	Handicap	-2.234*	-3.412**	-2.599**	-5.175**	-0.980	-0.988	0

*Correlational difference is significant at $p < 0.05$

** Correlational difference is significant at $p < 0.01$

DISCUSSION

Based on our results, we conclude that some order effects do take place when administering the OHIP. These findings can raise the question as to how these effects occur and to what extent these effects have implications in the clinical sense and for research efforts.

As has been noted in the 'Introduction', the OHIP is based on Locker's conceptual model of Oral Health, which makes the underlying structure of this instrument inherently hierarchical. Therefore, the relationships between the items have a part-whole combination. As for order effects, the first four scales show practically no response differences when changing item order and can therefore be considered as insensitive to order changes. This is probably the result of the content of the items in these subscales, which is more concrete and tangible. On the other hand results do show a heightened response on the last three subscales, when they are presented first. This suggests that when people are first confronted with questions about psychological and social enablement caused by dental problems, the impact on quality of life is regarded as greater than when confronted with these issues later in the questionnaire.

Accordingly, it is proposed that responses on items in the first four subscales influence the responses on items in the last three subscales in such a way that a consistency effect occurs. When considering the context of comparison, where the accessibility of information influences one's judgment, the information in the first subscales form a standard or norm on which one focuses (Sudman et al. 1996), in this case the occurrence of

oral symptoms or side effects. Therefore, the context of comparison is limited to specific oral impacts, such as oral pain. In a sense a person is made more conscious of how often certain aspects of oral health or disease have occurred, thus making the judgment on the previous items count for the interpretation of the less concrete and more general aspects of OHRQoL, which are represented by items contained by the social and psychological subscales (Weinberger et al. 2006).

So far the factorial structure of the OHIP has been neglected. Owing to a relatively small sample size, factor analyses could not be carried out. Nonetheless, the results show highly significant subscale intercorrelations, yet version 1 does show lower intercorrelations (indicating less overlap between subscales) than version 2. Ideally, correlations between subscales should be low and non-significant to be able to measure different constructs. This overlap may indicate the need for a more parsimonious model, with fewer dimensions, as supported by other studies into the factorial structure of the OHIP (John et al. 2004, John 2007, Mumcu et al. 2007). Furthermore, the intercorrelational differences between the subscales of both versions of the OHIP suggest variable factor structures as a result of order changes of the items.

Whether these order effects are clinically significant is a matter of perspective. There is no specific guideline for determining clinical significance, because small differences in mean scores can be statistically significant. Nevertheless, statistical significance is not equivalent to clinical significance (Crosby et al. 2003). To give some support to significance testing and to be able to interpret these outcomes, effect-size estimates can be used, but these estimates do not quantify clinical significance or relevance because that depends highly on the disease or the condition under consideration (Addy et al. 2005). In this study the largest effect size (0.19) was found on the Psychological disability scale. This indicates a small difference, but if one considers that this difference is only caused by an alternate order of items instead of a disease or treatment, it changes the perspective. Furthermore, both samples show a relatively high OHRQoL, compared with other sample studies (Slade 1997b, Allen et al. 1999). This is perhaps because of the sample population, which is relatively young and highly-educated and is therefore expected to have rather good oral health (Lopez et al. 2006). Above all, these characteristics have been shown to be indicative for lower OHIP scores and thus a higher OHRQoL (Slade 1997b). Therefore, it is possible that a more heterogeneous sample would have shown an even larger effect when it comes to changing the item order. A smaller effect is also possible; therefore, further research into this issue is needed.

However, for research efforts even small effects can have devastating impacts on outcomes and conclusions. The results indicate a consistency effect, which may be partly explained by the hierarchical underlying model of the OHIP. Additionally, subscale intercorrelational differences have been found, which could indicate different factor structures. For these reasons we recommend following the original item order of the OHIP, especially when considering the comparison of research outcomes with other studies. Furthermore, this study advocates more research into the factorial structure of the OHIP because high intercorrelations between subscales make the assumption of independent dimensions susceptible. In this study a selective sample was used, which makes generalizable conclusions precarious. However, the results do warrant more attention to be paid to order effects within the OHIP, or other OHRQoL questionnaires for that matter.

Acknowledgements- This study was supported by the Academic Centre for Dentistry Amsterdam (ACTA) and the Netherlands Institute for Dental Sciences (IOT).