CHAPTER 3
On the severity of impacts captured by the dimensions of the Oral Health Impact Profile

INTRODUCTION

The Oral Health Impact Profile (OHIP), developed by Slade and Spencer (Slade et al. 1994), is intended to evaluate the physical, psychological and social impacts of oral health on people's quality of life (Andersen et al. 1997, Oliveira et al. 2005). This instrument contains seven dimensions, which are based on Locker's conceptual model of oral health (Locker 1988, Slade et al. 1994). This model has its foundation in the Classification of Impairments, Disabilities and Handicaps developed by the World Health Organization (John et al. 2004). The dimensions are hierarchically ordered so that the impacts described by the dimensions are considered to be gradually more disruptive to one's life (Slade 1997a).

In Locker's model, impairment or disease may lead to Functional limitations, which in turn may lead to Physical pain and Psychological discomfort. Pain and discomfort can lead to disability (Physical, Psychological or Social), and disability can lead to Handicap (Slade et al. 1994). What can be inferred from this model is that it consists of an internal and an external component. The internal component focuses more on awareness of limitation or discomfort, and the external component is focused more on interpersonal and social experiences (Slade et al. 1994). As the OHIP is founded on this hierarchical model, one could argue that the complete set of events captured by dimensions higher in the hierarchy should be judged as having a more severe impact on quality of life than events described by lower dimensions.

Because Locker's hierarchical model has been proposed as the underlying basis of the OHIP, the aim of this study was to test the tenability of this model and to identify the relative severity of the impacts on daily life captured in the seven dimensions of the OHIP. It is hypothesized that dimensions situated higher in the hierarchy will be judged as having a more severe impact on daily life relative to dimensions lower in the hierarchy.

MATERIALS & METHODS

Participants were 235 psychology freshmen (71% female), for whom filling out questionnaires was a mandatory part of their course. 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 were given the appropriate information concerning the aim and general conclusions of this study.

The method of paired comparison, first introduced by Thurstone (Thurstone 1927), and the method of direct ranking were applied. One could use subscale mean scores, but subscale mean scores are, in fact, judgments made relative to an internal standard, whereas with the method of paired comparison all members of a pair are judged relative to all other members. With the method of direct ranking judgments are made simultaneously.

For the paired-comparison experiment the seven dimensions of OHIP, each constituting a problem which can occur in daily life, were captured in vignettes. These vignettes were used to run two balanced pairwise comparison experiments (example Fig.1). To adhere to the two components (internal and external) segmenting the hierarchical model of the OHIP, the dimensions were divided into two groups to form two pairwise comparison experiments. The first experiment was restricted to the dimensions: Functional limitations, Physical pain, Psychological discomfort, and Handicap. The second experiment was restricted to the dimensions Physical disability, Psychological disability, Social

disability, and Handicap. Both paired-comparison experiments, each consisting of four dimensions, gave ½n (n-1) = 6 pairs (Edwards 1957).

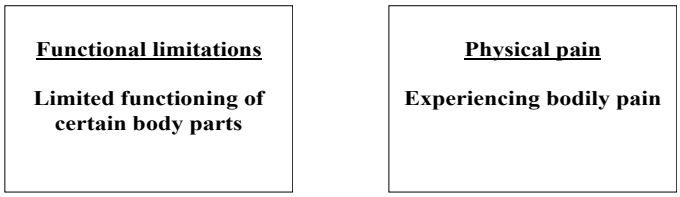

Fig. 1. An illustration of the vignettes

To obtain an optimum order for the presentation of pairs, Ross's method was used (Ross 1934). This method eradicates space and time errors, circumvents regular repetition, and maintains the greatest possible spacing between pairs involving any given member of the stimulus group (Ross 1934).

The intrasubject reliability was calculated by using Kendall's coefficient of consistence: ζ (zeta). It is defined as the consistency with which subjects make a paired judgment. Coefficient ζ can have a value ranging from 0 to 1, with a value of 1 indicating that a subject is perfectly consistent in his/her judgment Coefficient ζ is established by calculating the amount of circular triads. A circular triad is an inconsistency in ½n (n-1) judgments (Edwards 1957) and occurs, for instance, when a subject chooses A over B, B over C, and C over A, while the subject logically should have chosen A over C. The maximum amount of circular triads is equal to $(n^3-4n)/24$ for an even number of members. In this case there are two paired-comparison experiments with four members each, which give a maximum of two circular triads for each experiment. The more circular triads a subject has, the more inconsistency a subject displays and the lower the value of coefficient ζ is.

To determine the intersubject agreement, defined as the extent to which subjects agree in their comparative judgments amongst each other, Kendall's statistic *u* was calculated (Edwards 1957, Kendall et al. 1990). A value of 1 indicates perfect agreement and this value becomes smaller when agreement lessens.

The relative severity of the impact captured by the dimensions can be expressed in terms of scale values, as obtained by the Thurstone method (Edwards 1957). We refer to Edwards (Edwards 1957, p. 19-37), for the precise procedure used to calculate these scale values. The following is a summarisation of this procedure. To obtain scale values, the frequencies with which impact *I* is judged as more severe than impact *j* are placed in a frequency *(F)*-matrix, after which the proportion *(P)*-matrix is obtained by calculating the proportion corresponding to the frequencies in the *F*-matrix. Subsequently, the standardized *z*-values related to the proportions are ascertained to produce a *Z*-matrix. The z_{ij}-values in each column are then summed up and placed in order of succession, after which the mean *z*-values are calculated and regarded as the scale value. A constant is added to the scale values to make them positive. For convenience the largest negative scale value is used as a constant, forming a continuum of relative severity that has its origin in the

impact judged as least severe (with scale value zero). The scale values can be interpreted in terms of deviations relative to the mean of all other scale values.

In addition, the method of direct ranking was applied. This method allows all dimensions to be listed randomly after which subjects are asked to rank them. The order of presentation of the dimensions was the same for all subjects: Functional limitations, Physical disability, Handicap, Psychological discomfort, Social disability, Physical pain, and Psychological disability.

The two paired-comparison experiments and the direct ranking task were computerized and succeeded each other. Before starting the tasks, it was explained to the subjects that the questionnaire regarded problems in daily life. The pairs were presented one at a time, headed with the instruction: 'Which problem do you find most severe? Mark that problem with an X'. Subsequently, the direct ranking task was presented. This task was headed with the instruction: 'Rank the problems described below from 1 (most severe impact) to 7 (least severe impact)'.

RESULTS

The number of subjects related to the coefficient of consistency is shown in Table 1 for experiment 1 and 2. Nine subjects in experiment 1 and 22 subjects in experiment 2 were inconsistent in their judgment. Their data were excluded from further analyses. Furthermore, 49 out of 235 subjects failed to fill out the direct ranking task correctly. When comparing results with and without these 49 subjects differences were so minor that it was decided to report only the results of the smaller sample size (N=167).

Table 1. Number of circular triads (*d*), coefficient of consistency (ζ), and number of subjects

	d	ζ	Subjects
Experiment 1			
	0	1.00	226
	1	0.50	8
	2	0.00	1
Experiment 2			
	0	1.00	213
	1	0.50	20
	2	0.00	2

The results of the paired-comparison experiments 1 and 2 are presented in Tables 2 and 3, respectively. These tables show the frequency with which the column dimension was judged as having more impact on daily life than the row dimension. The rank of each dimension was determined by summing up the frequencies in each column.

The intersubject agreement was calculated and tested for significance using X^2 distribution, which represents the probability of finding a u value when judgments were made at random. The calculations resulted in u= 0.58, $X^2(6)$ = 570.77, p< 0.001 for experiment 1 and u= 0.73, $X^2(6)$ = 991.94, p< 0.001 for experiment 2. Therefore, both experiments showed highly significant agreement, beyond chance.

The results of experiment 1 (Table 2) show that subjects judged Handicap as having the most severe impact on daily life, followed by Psychological discomfort. The results of experiment 2 (Table 3) also show that subjects judge Handicap as having the most severe impact on daily life, followed by Psychological disability.

Table 2. The frequency with which each column dimension is judged as having more impact on daily living than each row dimension in experiment 1

		1	2	3	4
1	Functional limitations	-	41	89	152
2	Physical pain	126	-	117	130
3	Psychological discomfort	78	50	-	118
4	Handicap	15	37	49	-
	Total	219	128	255	400
	Rank	3	4	2	1

Table 3. The frequency with which each column dimension is judged as having more impact on daily living than each row dimension in experiment 2

		1	2	3	4
1	Physical disability	-	107	82	138
2	Psychological disability	60	-	62	92
3	Social disability	85	105	-	110
4	Handicap	29	75	57	-
	Total	174	287	201	340
	Rank	4	2	3	1

Scale values from experiment 1 and 2 are given in Table 4. According to the transformed scale values in experiment 1, the distance between the dimension with the least severe impact on daily living (Physical pain) and the dimension with the most severe impact on daily living (Handicap), was approximately 1 standard deviation relative to the mean scale value of all dimensions. In experiment 2, the distance between the dimension with the least severe impact on daily living (Physical disability) and the dimension with the most impact on daily living (Handicap), was less than one standard deviation relative to the mean scale value of all dimensions.

Table 4. Scale values of the dimensions in experiment 1 and 2

Dimensions	Scale value
Experiment 1	
Functional limitations	0.311
Physical pain	0.000
Psychological discomfort	0.511
Handicap	1.158
Experiment 2	
Physical disability	0.000
Psychological disability	0.460
Social disability	0.129
Handicap	0.688

As for the direct ranking task, the rank total of each problem was calculated. Results are presented in Table 5. Handicap and Psychological disability were judged as having the most severe impact on daily life, followed by Physical disability, Social disability, Functional limitations, Psychological discomfort, and Physical pain.

Table 5. Overall ordering according to the direct ranking task

Dimensions	Total	Rank
Handicap	475	1
Psychological disability	612	2
Physical disability	630	3
Social disability	717	4
Functional limitations	723	5
Psychological discomfort	741	6
Physical pain	778	7

DISCUSSION

In this study, two methods were applied to test the soundness of Locker's hierarchical model as the foundation of the OHIP. Furthermore, these methods were used, first to identify the relative severity of the impacts on daily life captured in the dimensions of the OHIP, and second to test the hypothesis that dimensions situated higher in the hierarchy will be judged as having a more severe impact on daily life relative to dimensions lower in the hierarchy.

Subjects' judgments on the paired-comparison experiments show Handicap being judged as having the most severe impact on daily life. This was expected because Handicap is situated highest in the hierarchy and can perhaps be seen as a combination of impacts described by the preceding dimensions in the model. Furthermore, in both paired-comparison experiments it seemed that psychological well-being was also regarded as an important attribute to daily life because the impacts of Psychological discomfort and Psychological disability were judged as having the second most severe impact on daily life. Although the relative distances indicated by the scale values were not particularly large (maximum approximately one standard deviation), they do however, show a difference in judging severity of impact on daily life. Moreover, given the selective nature of the sample, where subjects are expected to be relatively healthy, it could be that relative distances between severity judgments are higher in populations where disease and impairment are more prevalent.

The results of the direct ranking task showed a similar order, as was intended by Slade and Spencer (Slade et al. 1994). Because the dimensions were not divided into an internal and external component, as with the paired-comparison experiments, the results were difficult to compare. However, consistent with the findings of the paired-comparison experiments, Handicap and Psychological disability were also judged as having the most severe impact on daily life. In Contrast, Psychological discomfort was ranked relatively low, while Functional limitation was ranked relatively high, in the direct ranking task. Nevertheless, these contrasting results were merely found within the realm of the two components previously mentioned.

Most importantly, both paired-comparison experiments and direct ranking task, demonstrated the tenability of Locker's model as the foundation of the OHIP. The results of other studies support these findings (Nuttall et al. 2006, Baker 2007), but these studies based their support on frequency of impact, in other words, impacts that have actually been experienced and thus rest on an internal standard. Because, in this study, a relatively young and highly educated sample was used, which is expected to have a rather good oral health (Slade 1997b, Lopez et al. 2006), one could argue that this population does not have the

experience of certain oral impacts on daily life. For that reason, the judged severity of impact was used to explore the underlying model of the OHIP. Moreover, similar findings from different studies, using different methods, lend further support to these findings (Campbell et al. 1959).

Both paired-comparison experiments and direct ranking task indicate a difference between the severities of the impacts on daily life captured in the seven dimensions, which suggests that dimensions contribute differently to oral health-related quality of life (OHRQoL). The contribution of an item to a subscale has already been accounted for by including item weights (Slade et al. 1994), even though empirical evidence shows no overall improvement in the ability of the OHIP to discriminate between groups or to predict treatment needs when including item weights (Allen et al. 1997). This also applies to other measures, for that matter (McGrath et al. 2004). Perhaps this is a result of the fact that no apparent hierarchy exists within subscales, but only between subscales.

We therefore propose that when computing overall scores, the weight or contribution of a subscale should be accounted for. Given that overall scores and subscale scores are usually calculated by summing item scores, whereby the contribution of a subscale score to an overall score is a reflection of the number of items in the subscale, one could argue that dimensions, which are considered as having a more severe impact on the quality of life, should be represented by more items.

Whereas, within dimensions of the OHIP the relative differences between impacts of events have been compensated for by including item weights (Slade et al. 1994), no attempt has been made to compensate for the relative differences between dimensions. The OHIP does not include subscale weights, and although the number of items per subscale varies, the dimensions (Handicap and Psychological disability) that are, in fact, regarded by subjects as having the most severe impact on daily life, are reflected in the OHIP by the smallest number of items (six).

Although the limitations of this study are apparent in that a student sample was used, this does not weaken the fact that the relative contribution of subscales warrants more attention if one wants to unravel the true merit of OHRQoL. A complete paired-comparison experiment with all dimensions should be run amongst a more representative sample. The vignettes describing the impacts could be adjusted to accommodate more varied, less educated samples. We suggest that the OHIP could be modified by including subscale weights and/or by increasing the number of items, as suggested above.

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