DOI: 10.1111/j.1365-263X.2010.01066.x

INTERNATIONAL JOURNAL OF PAEDIATRIC DENTISTRY

Differential item functioning related to ethnicity in an oral health-related quality of life measure

JEFFERSON TRAEBERT[1], LYNDIE A. FOSTER PAGE[2], W. MURRAY THOMSON[3] & DAVID LOCKER[4]

[1]Public Dental Health Research Group, Universidade do Sul de Santa Catarina, Tubarão, Brazil, [2]Department of Oral Rehabilitation, School of Dentistry, University of Otago, Dunedin, New Zealand, [3]Department of Oral Sciences, School of Dentistry, University of Otago, Dunedin, New Zealand, and [4]Community Dental Health Services Research Unit, Faculty of Dentistry, University of Toronto, Toronto, Canada

International Journal of Paediatric Dentistry 2010; 20: 435–441

Objective. To assess whether an oral health-related quality of life (OHRQoL)measure showed differential item functioning (DIF) by ethnicity.
Methods. A simple random sample of 12- and 13-year-old schoolchildren enrolled in the Taranaki District Health Board's school dental service, New Zealand. Each child (n = 430) completed the Child Perception Questionnaire (CPQ_{11-14}) in the dental clinic waiting room, prior to a dental examination. The dataset included age, gender, ethnicity, and deprivation status. The general principle of the analytic plan was that equal scores from each CPQ_{11-14} item were expected from both non-Māori and Māori groups regardless of their ethnic group. Ordinal logistic regression was performed. The dependent variables were the CPQ_{11-14} items. The ethnicity group and each CPQ_{11-14} domain score were the independent variables. Non-uniform DIF was assessed through adding an interaction term for each CPQ_{11-14} sub-scale.
Results. Non-uniform DIF was found in two items, one in the Functional Limitations sub-scale and another in the Social Well-being sub-scale. Uniform DIF was found in one item of the Emotional Well-being sub-scale.
Conclusion. Both non-uniform and uniform DIF by ethnicity was found in three of 37 items of the CPQ_{11-14} questionnaire, showing it is important to perform DIF analysis when applying OHRQoL measures.

Introduction

When using a multi-item scale, it is assumed that the scale provides a correct description of the level of the underlying or latent construct it is supposed to assess. Constructs are described as such because they are observed not directly, but only indirectly through responses to a set of indicator items. Some examples of constructs are satisfaction with work, satisfaction with health status, anxiety, depression, and quality of life.

Because there is no single gold standard criterion measure of a given construct, measurement often proceeds by administering a series of questions (items), each of which presumably reflects in some way the underlying construct of interest. Establishing measurement equivalence across groups differing in such characteristics as education, gender, ethnicity, and others is important in health assessment, because researchers can then be confident in using the measure in different groups. Examination of differential item functioning (DIF) is central to the investigation of the equivalence of items contained in new and already established measures[1].

DIF analysis has been used in educational testing to investigate whether some items in a test are more difficult for a particular subgroup than for another, even if the participants have the same abilities or if they are at the same level of the underlying construct. For DIF to be deemed to be absent for a multi-item scale, all subjects at a given level of the construct measured should have the same probability of answering an item in the

Correspondence to:
Jefferson Traebert, Rua Dr. Armínio Tavares, 111/302, 88015-250 Florianópolis, SC, Brazil.
E-mail: jefferson.traebert@unisul.br

same way, regardless of their group membership[2].

A DIF study should address the different forms in which this can be observed. Uniform DIF refers to whether or not the difference in item response probabilities is constant across the scale: the item favours one group over another along the complete construct continuum. In non-uniform DIF the direction of the DIF differs along the construct scale and is tested for by adding an interaction term defined as the interaction between the construct level and group membership[3]. When both uniform and non-uniform DIF occur at the same time, it is called non-uniform asymmetrical DIF[4].

Increasing attention has been paid to the detection of DIF in health-related quality of life measures. It is an ongoing and integrated part of the analysis of multi-item scales in a study. As in any patient-related outcome measure, the observation of DIF in oral Health-related Quality of Life (ORHQoL) questionnaires is critical for judging its strength for measuring health outcomes in different groups and populations. As there is no guarantee that item behaviours will not be invariant across groups with different characteristics, taking values obtained from one group and using them in other group settings may lead to misleading findings[5].

DIF in multi-item scales distorts measurement. So it is fundamental to check whether item functions similarly across groups of people with different characteristics. This feature would allow researchers to investigate variations in instrument performance that are due to reasons other than those in the underlying construct. Failure to evaluate whether scales have DIF may pose a serious threat to the validity of between-group comparisons.

It has been shown that although Mäori have similar aspirations as all New Zealanders with regard to their quality of life (QoL), they also have certain unique views and aspirations which are specific to their culture and values, and any work done in the area of QoL for Mäori should capture both of these dimensions[6].

The aim of this paper was to determine whether an OHRQoL measure, namely the Children Perception Questionnaire ($CPQ_{11\text{-}14}$) had DIF related to ethnicity in a population-based sample of 12- and 13-year-old schoolchildren in New Zealand.

Methods

Study population

Detailed methods have been published elsewhere[7]. In brief, the study consisted of a simple random sample of 430 12- and 13-year-old schoolchildren enrolled in the Taranaki District Health Board's school dental service, New Zealand. Consent was obtained from parents and children before proceeding and ethical approval was obtained from the Taranaki Ethics Committee.

Study design

Each child completed the $CPQ_{11\text{-}14}$ questionnaire in the dental clinic waiting room, prior to dental examination. Questions asked about the frequency of events during the previous 3 months regarding four sub-scales: Oral Symptoms (six items), Functional Limitations (nine items), Emotional Well-being (nine items) and Social Well-being (13 items). The response options were: never (0), once/twice (1), sometimes (2), often (3) and every day/almost every day (4)[8].

The dataset also included gender, age, ethnicity (non-Mäori and Mäori), and deprivation level. The deprivation level was measured using an area based census measure, the NZDep2001 Index of Deprivation[9]. The area-based measure combines nine variables from the 2001 Census, which reflect aspects of material deprivation, and categorises each Census mesh block. This results in each mesh-block receiving a score that can range from 1 (lowest deprivation) to 10 (highest deprivation). For the analysis, address information was geocoded in order to enable each child to be allocated to a Census mesh block; this enabled allocation of each adolescent to an NZDep2001 score based on the area where he/she resided. Area with scores 1–3 were classified as 'low SES'; and those with scores 8–10 were classified as 'high SES'.

Data analysis

The general principle of the analytic plan was that equal scores in each CPQ_{11-14} item were expected from both ethnicity groups: non-Mäori and Mäori, after controlling for deprivation status (high, medium, low). The null hypothesis in each comparison was that there was no association between responses to the CPQ_{11-14} items and ethnicity group after controlling for an estimate of the construct measured by the sub-scale of which the item was a part. An estimate of the construct was obtained by summing the responses to the items in each subscale.

Since the CPQ_{11-14} items are scored on an ordinal scale, the analytic procedure used was ordinal logistic regression[10,11](all data were processed by SPSS software 15.0; SPSS Inc., Chicago, IL, USA). Given the distribution of responses to individual items, with responses to the lower scored categories being more probable, the negative log–log link function was employed.

The analytic procedure was proposed by Petersen *et al.*[12] For each item, the response to the item was first modelled as a logit-linear function of a dichotomous variable denoting ethnicity group, the sub-scale score of which the item is a part, and an interaction term which was the product of ethnicity group and the sub-scale score. If the interaction term is significant this provides evidence of non-uniform DIF. The interaction term was then removed and the analysis repeated with the remaining two variables in the model, ethnicity group and sub-scale score. Moderate-to-large uniform DIF was considered to be present if the ethnicity group variable was significant and the regression coefficient was numerically larger than 0.64; that is, the odds ratio was outside the interval 0.53–1.89. This standard was supplied by the Educational Testing Service[13] as adapted by Bjorner *et al.*[14] In the analysis, the non-Mäori group was coded 0 and the Mäori group was coded 1. This meant the Mäori group was the reference category. Consequently, when the regression coefficient for ethnicity group was positive, the non-Mäori respondents had a higher score on the item in question, meaning they reported more frequent impacts. When it was negative, the non-Mäori respondents had a lower score on the item and reported less frequent impacts.

Since the analysis of items within each sub-scale involved multiple comparisons, the *P*-value was adjusted to account for the number of analyses per sub-scale. Since the Oral Symptoms subscale has six items, a *P*-value was considered to be significant if it was less than 0.05/12 (six items and a test for non-uniform and uniform DIF for each), or 0.004. For the other sub-scales, the *P*-values were set as 0.003 for the Functional Limitation (nine items); 0.003 for Emotional Well-being (nine items) and 0.002 for Social Well-being (13 items).

Results

Characteristics of the sample

The study participation rate was 74.1%. Males slightly outnumbered females, and one in five participants was Mäori (Table 1). Scores on the CPQ_{11-14} ranged from 0 to 110 (Table 2).

DIF analysis

Table 3 summarises the results of DIF analyses in relation to ethnicity group, controlled by level of deprivation. Non-uniform DIF was

Table 1. Socio-demographic characteristics of a sample of schoolchildren (*n* = 430).

Variable	*n*	%
Gender		
Male	228	53.0
Female	202	47.0
*Age (years)**		
12	121	28.1
13	305	71.9
Ethnicity		
Mäori	88	20.5
Non-Mäori	342	79.5
*Level of deprivation**		
High	117	27.2
Medium	175	41.6
Low	120	27.9
Total	430	100.0

*Missing information.

Table 2. Descriptive data on the CPQ_{11-14}, domains, and ethnicity.

CPQ_{11-14}	Mean score (SD)	Range of observed scores
All		
Oral symptoms domain	4.9 (3.2)	0–19
Functional Limitation domain	5.5 (4.5)	0–30
Emotional Well-being domain	3.7 (5.0)	0–36
Social Well-being domain	4.0 (5.4)	0–30
Overall	18.1 (14.2)	0–110
Mäori		
Oral symptoms domain	5.4 (3.8)	0–18
Functional Limitation domain	5.1 (4.9)	0–30
Emotional Well-being domain	4.2 (6.2)	0–36
Social Well-being domain	4.5 (5.8)	0–29
Overall	19.1 (17.7)	0–110
Non-Mäori		
Oral symptoms domain	4.8 (3.0)	0–19
Functional Limitation domain	5.6 (4.4)	0–22
Emotional Well-being domain	3.5 (4.7)	0–31
Social Well-being domain	3.9 (5.3)	0–32
Overall	17.9 (14.3)	0–87

found in two items: 'breathed through your mouth' ($P < 0.001$) in the Functional Limitation sub-scale and 'avoided smiling or laughing when around other children' ($P = 0.002$) in the Social Well-being sub-scale. The item which met the criteria for moderate to large DIF ($P < 0.001$) was 'worried that you are not as good looking as others' in the Emotional Well-being sub-scale.

Discussion

Even though DIF analysis is not an end itself in most research using multi-item scales, it is relevant to carry it out before any group comparison. When comparing findings from OHRQoL questionnaires from different groups, it is fundamental to be confident that items are measuring the underlying construct in the same way. DIF analysis helps the researcher to be reasonably confident that item behaviour will not be invariant across groups with different characteristics. Otherwise, systematic error can be present and the study's findings can be adversely affected.

The CPQ_{11-14} performed differently in subgroups defined by ethnicity in three items. Two of them showed non-uniform DIF, one in the Functional Limitations sub-scale (breathed through your mouth) and one in the Social Well-being sub-scale (avoided smiling or laughing when around other children). In just one case, uniform DIF (worried that you are not as good-looking as others – Emotional Well-being) was found with a positive β coefficient (that is, a higher score), meaning non-Mäori respondents reported more frequent impacts.

Findings of significant DIF may have several causes. As for any statistical test, finding of significant DIF may arise purely by chance. If the DIF is not due to random variation, there still the possibility that it could reflect confounding by another variable. However, approaches to DIF analysis which allow for multivariate analysis (such as logistic regression), reduce the risk of misinterpretation due to confounding from other variables[2,12]. Nevertheless, the finding of significant DIF may be real. In this study, anthropologic and cultural reasons could hypothetically explain some differences in item behaviour.

For Mäori, QoL is linked to those aspects of well-being that are related to both Mäori culture and Mäori perspectives. The Mäori health perspective identifies wairua and whānau (spiritual and family), as cornerstones of health and well-being, as equally as important as physical and mental dimensions, and not only the well-being of Mäori individuals but the well-being of family. Family well-being is described as a Mäori-specific outcome because family well-being includes the use of Mäori values, customs and culture (such as language)[6]. As part of this, the sense that they do not want to draw attention to themselves in a predominantly New Zealand European setting may also account for the differences. Unfortunately, we did not measure constructs that reflect their values and beliefs so cannot expand on these in this group. However, this could be an interesting area for further research.

A limitation of this study could be the use of each CPQ_{11-14} domain scale score as the control variable. Ideally, a gold standard for the domain or item which was independent of the item responses would be better; however, such an ideal situation is rarely possible in the OHRQoL field. Thus, the scale score in each domain was the best choice. Using the total score for a set of items as a conditioning

Table 3. Results of uniform and non-uniform DIF analysis related to ethnicity (Mäori; non-Mäori) responses to CPQ_{11-14} in New Zealand.

	Non-uniform DIF	Uniform DIF		
CPQ_{11-14} items	***P*-value**	***β***	**95% CI**	***P*-value**
Oral symptoms				
Pain in your teeth, lips, jaws, or mouth	0.250	0.405	0.136/0.946	0.142
Bleeding gums	0.625	−0.150	0.632/0.331	0.541
Sores in your mouth	0.134	−0.075	−0.571/0.420	0.765
Bad breath	0.087	0.115	−0.345/0.576	0.623
Food stuck in or between your teeth	0.062	0.236	−0.377/0.549	0.717
Food stuck in the top of your mouth	0.971	−0.219	−0.750/0.312	0.419
Functional Limitations				
Breathed through your mouth	**<0.001**	0.396	−0.036/0.828	0.073
Taken longer than other to eat a meal	0.997	0.158	−0.313/0.629	0.510
Had trouble sleeping	0.497	−0.363	−0.917/0.191	0.199
Difficult to bite or chew food like apples, corn on the cob or steak	0.718	0.327	−0.272/0.927	0.284
Difficult to open your mouth wide	0.301	−0.508	−1.191/0.175	0.145
Difficult to say any words	0.626	−0.586	−1.208/−0.037	0.065
Difficult to eat foods you would like to eat	0.798	−0.304	−0.957/0.349	0.362
Difficult to drink with a straw	0.752	−0.478	−1.816/0.859	0.483
Difficult to drink or eat hot or cold foods	0.206	0.263	−0.288/0.813	0.350
Emotional Well-being				
Felt irritable or frustrated	0.051	0.280	−0.281/0.841	0.327
Felt unsure of yourself	0.323	0.394	−0.95/1.083	0.263
Felt shy or embarrassed	0.284	−0.510	−1.056/0.037	0.067
Been concerned what other people think about your teeth, lips, mouth or jaws	0.251	0.550	−0.080/1.181	0.087
Worried that you are not as good-looking as others	0.191	**2.010**	**1.073/2.946**	**< 0.001**
Been upset	0.561	−0.898	−1.564/−0.232	0.008
Felt nervous or afraid	0.014	−0.205	−0.786/0.377	0.490
Worried what you are not as health as others	0.289	−0.226	−0.789/0.336	0.430
Worried that you are different than other people	0.710	−0.137	0.842/0.568/0.483	0.703
Social Well-being				
Missed school because of pain, appointments, or surgery	0.155	0.523	−0.119/1.164	0.110
Had a hard time paying attention in school	0.786	−0.409	−1.029/0.211	0.196
Had difficulty doing your homework	0.945	−0.353	−0.977/0.272	0.269
Not wanted to speak or read out loud in class	0.374	−0.647	−1.258/−0.036	0.038
Avoid taking part in activities like sports, clubs, drama, music, school trips	0.689	−0.243	−0.966/0.480	0.510
Not wanted to talk to other children	0.546	0.263	−0.600/1.127	0.293
Avoided smiling or laughing when around other children	**0.002**	0.762	0.029/1.496	0.042
Had difficult playing a musical instrument such as a recorder, flute, clarinet, trumpet	0.051	−1.118	−1.961/−0.466	0.009
Not wanted to spend time with other children	0.237	−0.034	−0.839/0.772	0.935
Argued with other children or your family	0.534	−0.115	−0.611/0.380	0.649
Other children teased you or called you names	0.009	1.064	0.329/1.798	0.005
Other children made you feel left out	0.139	0.091	−0.630/0.813	0.804
Other children asked you questions about your teeth, lips, jaws or mouth	0.142	0.528	−1.062/0.233	0.111

A positive β reflects a higher item score for non-Mäori than for Mäori at a given CPQ_{11-14} domain score.
Statistical significance in bold.
Controlled by socio-economic status.

measure, those items with DIF can lead to the total score being a poor estimator of the underlying construct[1]; however, this is probably not the case in this study because the three items found with DIF were in different sub-scales.

Although it is unusual to include DIF analyses in the reporting of OHRQoL outcomes, it is important to do so. The question could be similar to whether one should include possible confounding variables in epidemiological analyses[2]. Groenvold and Petersen[2] have pointed out that carrying out DIF analysis and interpreting the results takes time and effort. In a realistic way, they suggest that DIF analysis is of particular importance: (1) in analyses of great clinical importance; (2) if the unidimensionality of the scale is questionable; (3) if DIF has previously been detected in the scale in relation to the variable; (4) in analyses focusing on variables which may be associated with DIF, such as ethnicity gender and age; and (5) in studies using questionnaire versions in more than one language.

The researcher should try to understand whether significant DIF found with a given item will significantly affect the findings. One possibility is to remove the item with DIF from the multi-item scale and to compare the results from this DIF-free scale with those from the full scale. Alternatively, analysis at single-item level could be performed and compared[2].

However, it is important to consider whether or not it is appropriate to include DIF assessment in a particular study. Different constructs may be more or less affected by DIF in respect of demographic factors. DIF related to educational profile can be more important than that related to gender in constructs such as cognitive ability, whereas gender-related DIF to can be more important in constructs related to physical functioning and abilities[1]. Nevertheless, OHRQoL measures are particularly susceptible to sociodemographic characteristics, including gender, age and ethnicity, according to the cultural contexts in which they are applied. In this way, it is important to consider DIF analysis even if cross-cultural adaptations are rigorously performed according to standard international recommended guidelines.

More research should be done in order to assess the impact of DIF on findings using OHRQoL measures. This includes the understanding about if found DIF are of clinical or epidemiological importance, roles played by uniform, non-uniform and non-uniform asymmetrical DIF and approaches for elimination of DIF in OHRQoL measures.

Conclusion

Both uniform and non-uniform DIF were found in three of 37 items of the $CPQ_{11\text{-}14}$ questionnaire related to ethnicity in New Zealand schoolchildren. It is relevant to perform DIF analysis when applying OHRQoL measures.

What this paper adds

- This paper undertakes differential item function analysis in assessing the measurement equivalence of items in an oral health outcome questionnaire.
- The study found that three items of $CPQ_{11\text{-}14}$ were not equivalent when applied for Mäori and non-Mäori children. Differences detected in those items may be an artefact of the measurement process rather than a reflection of actual sub-group differences.

Why this paper is important to paediatric dentists

- Questionnaires such as $CPQ_{11\text{-}14}$ have good potential for application in clinical practice in order to get information on functional and psychosocial aspects affecting patients. However, paediatric dentists should be aware that scores derived from these scales can function in different ways in different sociodemographic groups (and especially different ethnic groups).

Acknowledgements

When undertaking this research, J. Traebert was supported by a post-doctoral scholarship by CAPES, Ministry of Education, Brazilian Federal Government. The authors are grateful to the New Zealand Dental Association Research Foundation and the Taranaki District Health Board for funding the New Zealand data collection, and to the Iwi of Taranaki for their support.

References

1 Teresi JA, Fleishman JA. Differential item functioning and health assessment. *Qual Life Res* 2007; **16**: 33–42.

2 Groenvold M, Petersen MA. The role of use of differential item functioning (DIF) analysis of quality of life data from clinical trials. In: Fayers P, Hays R (eds) *Assessing Quality of Life in Clinical Trials*, 2nd

edn. Oxford: Oxford University Press, 2005: pp. 195–208.
3 Crane PK, Gibbons LE, Ocepek-Welikson K *et al.* A comparison of three sets of criteria for determining the presence of differential item functioning using ordinal logistic regression. *Qual Life Res* 2007; **16**: 69–84.
4 Hidalgo MD, Gomez J. Nonuniform DIF detection using discriminant logistic analysis and multinomial logistic regression: a comparision for polytomous items. *Qual Quant* 2006; **40**: 805–823.
5 Badia X, Roset M, Herdman M. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Med Decis Making* 2001; **21**: 7–16.
6 Kooyela V. Quality of Life Indicators for Mäori: A discussion document for the Mäori potential forecast report. A report prepared for Te Puni Kokiri the Ministry of Mäori Development, New Zealand, 2007.
7 Foster Page LA, Thomson WM, Jokovic A, Locker D. Validation of the Child Perceptions Questionnaire ($CPQ_{11\text{-}14}$). *J Dent Res* 2005; **84**: 649–652.
8 Jokovic A, Locker D, Stephens M, Kenny D, Tompson B, Guyatt G. Validity and reliability of a questionnaire for measuring child oral-health-related quality of life. *J Dent Res* 2002; **81**: 459–463.
9 Crampton P, Salmond C, Kirkpatrick R. *Degrees of Deprivation in New Zealand: An Atlas of Socioeconomic Difference*, 2nd edn. Auckland: David Bateman Ltd, 2004.
10 Swaminathan H, Rogers HJ. Detecting differential item functioning using logistic regression procedures. *J Educ Meas* 1990; **27**: 361–370.
11 French AW, Miller TR. Logistic regression and its use in detecting differential item functioning in polytomous items. *J Educ Meas* 1996; **33**: 315–332.
12 Petersen MA, Groenvold M, Bjorner JB *et al.* Use of differential item functioning to assess the equivalence of translations of a questionnaire. *Qual Life Res* 2003; **12**: 373–385.
13 Zieky M. Practical questions in the use of DIF statistics in test development. In: Holland PW, Wainer H (eds). *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1993: pp. 93–106.
14 Bjorner J, Kreiner S, Ware J, Damsgaard M, Bech P. Differential item functioning in the Danish translation of the SF-36. *J Clin Epidemiol* 1998; **51**: 1189–1202.