Abstract

Introduction: Oral Health Impact Profile (OHIP) is a widely used psychometric instrument or scale developed in English to measure Oral Health-related Quality of Life (OHQoL), and there has been many translations of the instrument into other languages, including Korean.

Purpose: My thesis examines the validity and cultural equivalence of the English and Korean versions of the scale by answering the questions: “What methods are available to validate the cultural equivalence of psychometric instruments?” and “How culturally appropriate and valid is the Korean version of the short-form of the OHIP (OHIP-14K)?

Method: Ten Korean dental experts fluent in English and Korean independently assessed the clarity, relevance, and cultural equivalence of the OHIP-14K and offered suggestions for improving the cultural sensitivity and validity of the instrument content. The item-level Content Validity Index (I-CVI) was used to measure the validity of each item from the experts’ ratings followed by the calculation of Scale-level Content Validity Index (S-CVI) as the proportion of content valid items. Additional analyses including the average deviation index (AD_M) and Kappa statistics (K_free) were performed with the clarity index (CI), relevance index (RI) and cultural equivalence index (CEI) to measure the level of agreement between the experts.

Results: The experts rated the OHIP-14K as mostly clear (S-CVI= 0.93), but they were concerned about the relevance of many items to the expected domains of the instrument (S-CVI = 0.42) and about its cultural equivalence (S-CVI = 0.50) to the English version. However, there was much disagreement between the experts as measured by the RI (K_free = 0.19 to 1.00) and CEI (AD_M = 0.36 to 0.96).

Conclusion: The relevance and cultural equivalence of the OHIP-14K to the original English version of the OHIP-14 are not strong. Suggestions are offered for improving the OHIP-14K, which needs further testing within the Korean populations.
Preface

Ethics approval for the present research was obtained from the Research Ethics Board of the University of British Columbia (BREB#: H10-01023).

I, Jaesung Seo, claim sole authorship of the content of the present thesis.
Table of Contents

Abstract .................................................................................................................................................. ii
Preface .................................................................................................................................................. iii
Table of Contents .................................................................................................................................. iv
List of Tables ......................................................................................................................................... viii
List of Figures ....................................................................................................................................... ix
Glossary ................................................................................................................................................ x
Acknowledgements ............................................................................................................................. xi
Dedication ............................................................................................................................................ xiii

Chapter 1: Introduction ......................................................................................................................... 1

1.1 Objectives and Research Questions .............................................................................................. 4

Chapter 2: First Component – Literature Review ................................................................................ 6

2.1 Theoretical Frameworks ................................................................................................................. 6

2.1.1 Biomedical Model ....................................................................................................................... 7
2.1.2 Sociomedical Model .................................................................................................................... 8

2.1.2.1 Development of the Original Forms of the Oral Health Impact Profile ......................... 9
2.1.2.2 Criticisms of Locker’s Model and OHIP ................................................................................ 13

2.1.3 Biopsychosocial Model ............................................................................................................. 14
2.1.4 OHQoL Measurement in Korean ............................................................................................... 16

2.2 Validity Theory ............................................................................................................................... 17

2.2.1 Trinitarian Concepts of Validity and OHIP-14K ........................................................................ 17
2.2.2 Unitarian Concept of Validity .............................................................. 20

2.3 Cross-Cultural Equivalence Theory ......................................................... 21
  2.3.1 Method Equivalence ........................................................................... 23
    2.3.1.1 Scale Bias ................................................................................... 23
    2.3.1.2 Administration Bias ..................................................................... 25
  2.3.2 Item Equivalence ................................................................................. 27
  2.3.3 Construct Equivalence ......................................................................... 28

Chapter 3: Systematic Synthesis of the Literature ............................................. 31

  3.1 Search Strategy ....................................................................................... 31
    3.1.1 Forward Translation ......................................................................... 34
    3.1.2 Back-Translation .............................................................................. 35
    3.1.3 Committee Translation ..................................................................... 36
    3.1.4 Pilot Testing ...................................................................................... 37
    3.1.5 Statistical Validation ........................................................................ 37
  3.2 Content Validity ....................................................................................... 40
  3.3 Cross-cultural Equivalence ...................................................................... 41

Chapter 4: Second Component – Empirical Exercise Methods .......................... 45

  4.1 Selection of Bilingual Subject Matter Experts ......................................... 46
  4.2 Development of Questionnaire on Content Validity ................................. 49
    4.2.1 Clarity Index .................................................................................... 50
    4.2.2 Cultural Equivalence Index ............................................................... 50
    4.2.3 Relevance Index .............................................................................. 51
  4.3 Administration of Content Validation Questionnaire and Data Collection ... 52
  4.4 Data Analysis ........................................................................................ 52
Appendix A........................................................................................................................................ 134
   A.1 Coding Search for Inclusion in Systematic Synthesis ......................................................... 134
Appendix B ........................................................................................................................................ 135
   B.1 English Version of Oral Health Impact Profile-14............................................................... 135
   B.2 Oral Health Impact Profile-14K (Bae et al., 2007) ............................................................. 137
   B.3 Oral Health Impact Profile-14K (Kim & Min, 2009) ............................................................. 139
   B.4 Oral Health Impact Profile-14K (Kim & Min, 2008) ............................................................. 141
Appendix C ........................................................................................................................................ 143
   C.1 Computation of Item-level AD$_M$ for Clarity and Cultural Equivalence Index............. 143
   C.2 Computation of AD$_M$ for Clarity and Equivalence Index .............................................. 143
   C.3 Computation of Multirater $K_{free}$ for Relevance index items........................................ 144
   C.4 Interpretation of Multirater $K_{free}$ Statistics ................................................................. 144
Appendix D ........................................................................................................................................ 145
   D.1 Informed Consent.................................................................................................................... 145
   D.2 Cover Letter and Content Validation Questionnaire.......................................................... 149
List of Tables

Table 1. Purposes for which the Oral Health Impact Profile has been applied. .........................2
Table 2. The original short form of the Oral Health Impact Profile (OHIP-14) questions and theoretical domains (Slade, 1997).................................................................12
Table 3. The various methods used to translate and culturally adapt the OHIP ..........................39
Table 4. Background information of the subject matter experts.................................................48
Table 5. Item-level content Validity Index (I-CVI) and Average Deviation from the Mean Index (ADM) of OHIP-14K instruction, response format, event frequency, and items in the Clarity and Equivalence Indices .........................................................................................57
Table 6. The SMEs’ classification of items into the seven OHIP domains, Item-level Content Validity Index (I-CVI), Free-Marginal Kappa (K_free), and levels of agreement on the Relevance Index........................................................................................................................60
Table 7. Suggested revisions of OHIP-14K................................................................................74
Table 8. Suggested revised version of OHIP-14K.......................................................................81
List of Figures

Figure 1. International Classification of Impairment, Disability and Handicap (WHO, 1980)..... 9
Figure 2. Locker's model of oral health (Locker, 1988) [Reproduced with the permission of the editor of Community Dental Health] ................................................................................................................................. 10
Figure 3. Various aspects of validity and equivalency for cross-cultural measurement and comparison of OHQoL ........................................................................................................................................ 30
Figure 4. Stages of OHIP-14K development and proposed subscale indices of the content validity questionnaire........................................................................................................................................ 49
Figure 5. Possible outcomes of content validation study of OHIP-14K........................................ 54
Figure 6. An existential model of oral health (MacEntee, 2006) [Reproduced with the permission of the editor of Community Dental Health]. ......................................................................................... 88
Figure 7. The atomic model of oral health (Brondani et al., 2007) [Reproduced with the permission of the editor of Gerodontology]........................................................................................................................................ 89
Glossary

Bias: “deviation from the truth” (Grimes & Schulz, 2002, p. 2).

Construct: a theoretical representation of a concept, trait, attribute, or any variable that cannot be directly measured (Messick, 1989). A construct (e.g., quality of life) is often multi-faceted and operationally defined by a theory.

Construct validity: how well a scale measures the construct of interest theorized by the underlying model (Hubley & Zumbo, 1996).

Content validity: the relevance and representativeness of the scale content (e.g., items or questions) to the construct of interest and its appropriateness for the target population (Beck & Gable, 2001).

Convergent validity: “the notion that the application of two or more different measures of the same trait should lead to highly correlated, or convergent, results” (Larcker & Lessig, 1980).

Criterion validity: empirical correlation between a test score and concurrent or predictive criteria (Messick, 1989).

Cultural adaptation: a process of translating and modifying an existing scale to reflect cultural values and norms of target populations for use in a new country or culture (Guillemin et al., 1993).

Discriminant validity: “the notion that the correlation between different measures of the same trait should be greater than the correlation between different measures of different traits and the same measure of different traits” (Larcker & Lessig, 1980).
**Equivalence**: unbiased measurement of actual differences among different cultural groups (Cella et al., 1996). Cross-cultural equivalence is broadly divided into measurement equivalence (the extent to which the item content is interpreted the same way across cultures) and structural equivalence (the degree of congruency in the underlying theoretical structure and relations of items measured by subscales) (Byrne & Watkins, 2003).

**Face validity**: respondents’ perception of the scale’s appropriateness for the purpose of measuring the construct of interest (Nevo, 1985).

**Fidelity**: the extent to which a culturally adapted version of a scale is faithful to the original scale (Corless et al., 2001).

**Measurement**: a process of linking the observable empirical indicants to the underlying unobservable *construct* (Zeller & Carmines, 1980).

**Reliability**: the extent to which repeated measurement of the same property produces the same result.

**Scale**: “a standardized procedure for sampling behaviour and describing it with categories or scores” (Gregory, 2007).

**Validity**: confidence that the scale measures what it purports to measure in the target culture or population (Sandelowski, 1986).

**Acknowledgements**

I offer my utmost thanks to my research supervisor Dr. Mario Brondani for his support in my personal, academic, and professional endeavours. I would also like to thank my co-supervisor
Dr. MacEntee and my MSc committee, Drs. Bryant, Graf and Wong, for their teachings, guidance and encouragement. Finally, I would like to thank my wife Ewa Seo and family members for their selfless devotion and unconditional support for turning my academic pursuit into reality.
Dedication

This thesis is dedicated to my wife, Ewa Seo, for always being there for me and enduring struggles and hardships together with me.
Chapter 1: Introduction

Emerging from the landmark study by Cohen and Jago on sociodental indicators (1976), oral health-related quality of life (OHQoL) has come to represent a psychological construct defined as self-reports from people pertaining to the functional, psychological and social impacts of oral problems on their quality of life (Gift & Redford, 1992). To date, more than a dozen OHQoL measurement tools have been developed, as summarized by Slade (1997), Allen (2003), Johns and colleagues (2004), Hebling and Pereira (2007), and again by Brondani and MacEntee (2007). Unlike clinical indices, their primary objective is to capture biopsychosocial consequences of oral problems that are perceived to be important by lay respondents (Allen, 2003). Among these is an internationally recognized and frequently used scale called the Oral Health Impact Profile (OHIP) (Slade & Spencer, 1994). Originally developed in Australia, OHIP is a self-report measure of dysfunction, discomfort and disability due to oral conditions. It consists of 49 questions representing seven domains (impairment, functional limitation, discomfort/pain, physical, psychological and social disabilities and handicap) and 5-point Likert response categories (4 = “very often,” 3 = “fairly often,” 2 = “occasionally,” 1 = “hardly ever”, and 0 = “never,” with an optional “don’t know”) (Slade, 1997). A shortened version of the OHIP called OHIP-14 was later developed based on the 14 of original 49 questions (subsets of 2 questions for each of the seven domains, Table 2). The short form is preferred to the longer OHIP-49 by many researchers and respondents because it takes less time to complete, costs less for administration and data management, and imposes less burden on the respondent, especially if they are frail, thus helping to improve the item-response rate (Saub et al., 2005). Both the long and short forms of the OHIP have been widely used as an evaluative and discriminative tool for research and clinical outcomes (Table 1).
Table 1. Purposes for which the Oral Health Impact Profile has been applied.

<table>
<thead>
<tr>
<th>Purpose of Application</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health care utilization, resource allocation and the quality of care and interventions</td>
<td>Atchison, 1997; Pearson et al., 2007; Walker &amp; Kiyak, 2007</td>
</tr>
<tr>
<td>Oral health promotion and disease prevention programs</td>
<td>Locker, 1995</td>
</tr>
<tr>
<td>Patient-centred care-planning</td>
<td>Slade, 1997; Jones et al., 2004</td>
</tr>
<tr>
<td>Patient outcome-evaluation</td>
<td>Gift, 1997; Hebling &amp; Pereira, 2007</td>
</tr>
<tr>
<td>Prioritization of research and treatment options</td>
<td>Cohen, 1997; Calabresse et al., 1999</td>
</tr>
<tr>
<td>Interdisciplinary health-care</td>
<td>Hayran et al., 2009</td>
</tr>
<tr>
<td>Measurement of psychological burden of oral problems</td>
<td>Cohen 1997; Calabrese &amp; Friedman, 1999; Sampogna et al., 2009</td>
</tr>
<tr>
<td>Patients’ assessment of care</td>
<td>Allen, 2003</td>
</tr>
<tr>
<td>Impact of oral disorders on patients with systemic diseases</td>
<td>Mumcu et al., 2006; Jeganathan et al., 2011</td>
</tr>
</tbody>
</table>

Although commonly applied across all age groups, OHQoL is particularly important for older adults who have a greater accumulation of chronic oral problems, and where the primary treatment goals are to improve basic functions (Kressin et al., 1996). The growing interest in OHQoL measurement has also spread widely, and the original OHIP in English has been translated into at least 19 languages (Appendix B.1), which has subsequently stimulated cross-cultural comparisons of OHQoL (Allison et al, 1999; Kawamura et al., 2000; Kawamura et al., 2001; Steele et al., 2004; Slade et al., 2005; Bae et al., 2007; Rener-Sitar et al., 2008).

This momentum in cross-cultural measurement of OHQoL has occurred also in South Korea, which has one of the fastest-aging populations (Cha, 2009). The National Statistics Office (2007) of South Korea projected that this country will become the most aged country globally by 2050, when 38.2% of the population will be aged 65 and older, nearly double the projected world average of 16.2%. Despite the drastic demographic shift, oral health is perceived as a low
priority by South Korean older adults and their caregivers. However, the assessment of OHQoL in this population could bring benefits in terms of detecting oral and nutritional problems (Jung & Shin, 2008) and treatment needs (Bae et al., 2007) by improving public oral health promotions (Ha et al., 2012) and studying oral health outcomes (Cha et al., 2011). The benefits of OHQoL cultural adaptation could be conferred not only to elders in Korea but also to Korean immigrants elsewhere who speak Korean. OHQoL measurement in Korean is important in the Canadian context as Canada is home to the world’s third-largest Korean immigrant population, whose settlement is concentrated in Vancouver and Toronto (Lee, 2010). They represent the seventh-largest incoming group of immigrants in Canada, and the third-largest visible minority group in the Greater Vancouver with approximately 2.3% of the total population (City of Vancouver, 2000). Moreover, 12.0% of them are aged over 55 years (Statistics Canada, 2006). The first generation of elderly Korean immigrants, who are not fluent in English, are most likely to benefit from OHQoL assessment in Korean.

Originally developed for use in Southern Australia, the OHIP has been culturally adapted to the Korean population by Bae and colleagues (2007). In lieu of developing an entirely new OHQoL scale (i.e., de novo method), adapting a supposedly validated scale to target cultures has the advantage of providing cheaper, quicker and easier scale preparation; cross-language support; inclusion of immigrants; and fair cross-cultural comparisons (Hambleton & Patsula, 1998; Guillemin et al., 1993). Cultural adaptation of the scale is usually followed by the extensive validation efforts. Moreover, given the vast amount of literature on cross-cultural comparisons of OHIP scores, the researchers worldwide might hold the view that the current methods of cultural adaptation and validation preserve the validity and equivalence of the original version of OHIP for obtaining valid and reliable information. However, it remains unclear how these properties are conceptualized in cross-cultural OHQoL researches and established by the current
standards of cultural adaptation and validation methods. More importantly, to what extent a translated version of the OHIP exhibits these properties has yet to be fully understood.

1.1 Objectives and Research Questions

Given these introductory remarks, the overriding research aim of my thesis was to describe the concepts of validity and equivalence in the context of cross-cultural OHQoL measurement and to recommend an effective process for investigating these properties. In regard to the aims of my study, I have organized my dissertation into two major components. The first component is more theoretical and is comprised of a synthesis of the existing literature that will allow me to critically examine the theoretical frameworks underpinning the existing scales of OHQoL with a main emphasis on the OHIP. I will then analyze the literature on different strategies for the adapting and validating English-based psychometric measures for other cultures. For this I will pose three research questions, which, in turn, will help me to discuss theories of validity, equivalence and measurement. The first two questions are:

1. How are validity and cultural equivalence conceptualized in cross-cultural settings?
2. What are the strengths and weaknesses of existing methods for culturally adapting and validating the OHIP-14?

The second component of my dissertation is more empirical by using exercises with SMEs to assess the content validity and cultural equivalence of the Korean version of OHIP-14 (OHIP-14K) with the original OHIP-14. In this section I will address my third research question:

3. To what extent does OHIP-14K exhibit content validity and cultural equivalency in Korean?
I will then discuss the methods I used and the results I gathered from this exercise. Finally, I will proceed with the concluding discussion and offer the implications and future directions of my findings.
Chapter 2: First Component – Literature Review

2.1 Theoretical Frameworks

The theoretical approach to validating the OHIP requires adherence to psychometric theory, which refers to “all those aspects of psychology which are concerned with psychological testing, both the methods of testing and the substantive findings” (Kline, 2000, p. 1). There are two main branches of measurement theory in the field of psychometrics: classical test theory (CTT) and item response theory (IRT). CTT assumes that a respondent has an observed score and a true score, and that these can be interpreted against a reference or norm group. IRT makes similar assumptions about the observed and true scores, but it delves deeper into modelling the relationship between responses to each item and each construct (Reeve & Masse, 2004), allowing for more advanced statistical analyses.

Both CTT and IRT are compatible with the concepts of validity and cultural equivalence, but my choice of CTT as the theoretical measurement framework is motivated by its alignment with the OHIP design. Many features of the OHIP are compatible with the assumptions of CTT in terms of equal contribution of individual items to the total score, constancy of the true score from test to retest, and dependence on a sample of participants to determine the properties of the scale, such as item discrimination (Magno, 2009; Wong et al., 2010). Moreover the OHIP, like most OHQoL measures, is validated with methods grounded in CTT, including internal consistency, criterion-related validity, and test-retest reliability (Locker, 2008).

IRT, on the other hand, requires unidimensionality of the construct as in the case of the construct of alcohol dependence. (Helzer et al., 2008). Therefore, CTT is a better alternative for
studying multidimensional constructs such as OHQoL and used as a guiding principle for my literature review and for the design and analysis of my validation study.

Now I will review the literature on the concept of OHQoL. Over the years, efforts to conceptualize and quantify OHQoL have been shaped by the evolving frameworks or models of health and disability.

2.1.1 Biomedical Model

Traditionally, the most predominant theoretical framework in oral health research had been the biomedical approach, which focuses primarily on diagnoses and pathologies when defining health (Levasseur et al., 2004). The biomedical view of health and illness was the result of a mechanical view of biology called the Cartesian mechanical paradigm, which constituted the foundation of modern medicine (Capra, 1982). From this biomedical point of view, an absence of oral health is synonymous with the clinical presence of disease and tissue destruction (Slade & Sanders, 2003), as exemplified by WHO’s definition of oral health as “a state of being free from chronic mouth and facial pain, oral and throat cancer, oral sores, birth defects such as cleft lip and palate, periodontal disease, tooth decay and tooth loss, and other diseases and disorders that affect the oral cavity” (2008, para. 1). Pursuant to this definition, oral health was frequently measured with clinical indices such as decayed, missing, filled teeth (DMFT) (Dean, 1942), periodontal index (Russell, 1956), community periodontal index of treatment needs (CPITN) (Cutress et al., 1982), gingival bleeding index and plaque index (Silness & Loe, 1964), and so on. However, the biomedical model can only define the presence or absence of oral problems; it cannot account for behavioural and subjective inner experiences prompted by the problems (Warner, 1999). Researchers have long been perplexed by the lack of correspondence between the clinical status and subjective impacts of oral health (Leao & Sheilham, 1995). The
traditional clinical measures are too narrow in scope to capture various dimensions of human experience (Slade & Locker, 1994) and they provide limited scope of the consequences of oral disorders (Reisine, 1985). In response to these criticisms and the need for global scales and broader conceptual frameworks, two main schools of thought have emerged: sociomedical and biopsychosocial models.

2.1.2 Sociomedical Model

From the sociomedical perspective, individuals’ oral health status is directly linked to their oral health’s effect on their task performance and usefulness to society, as exemplified by Dolan’s definition: “a comfortable and functional dentition which allows individuals to continue their desired social role” (1993, p. 37). The need for tools with which to measure these social factors and impacts of oral health, such as personal lifestyle or cultural and ecological factors, was first recognized by Cohen and Jago (1976) and has led to development of more than a dozen OHQoL measures (see Slade, 1997; John et al., 2004; MacEntee, 2006; Brondani & MacEntee, 2007; Hebling & Pereira, 2007). They attempt to describe lay people’s perception of their own oral health and, in tandem with clinical measures, to determine health care goals and assess the degree to which they are met (Cushing et al., 1985) (Table 1). Three interdependent sociomedical frameworks have been used as the theoretical basis for formulating OHQoL measures: Parsonian Sick Role Theory (Parsons, 1951); International Classification of Impairments, Disabilities and Handicaps (ICIDH) (WHO, 1980); and Locker’s model of oral health (Locker, 1988).

Resine (1981) first proposed the use of a sociomedical framework called Parson’s sick role theory, which posits that a sickness warrants exemption from fulfilling one’s social role; that it
is the role of health professionals to restore a sick person’s health; and that an illness is an undesirable and socially deviant behaviour, so sick people must seek treatment from a health professional in order to substantiate their sick roles (Parsons, 1951). The sick role theory first found its place in the WHO’s framework of disability, known as the ICIDH (WHO, 1980). ICIDH explains consequences of disease in three dimensions (Figure 1):

- Impairment: organ-system-level loss of structure or function.
- Disability: person-level loss of the ability to function in ways considered normal for a human being.
- Handicap: societal-level disadvantage for the person created by the intersection of the impairment or disability with the person’s environment and roles.

![Figure 1. International Classification of Impairment, Disability and Handicap (WHO, 1980)](image)

2.1.2.1 Development of the Original Forms of the Oral Health Impact Profile

Presently, the key concepts of ICIDH and the sick role theory provide the theoretical basis for the majority of OHQoL scales, including the OHIP (MacEntee, 2006). For instance, the theoretical framework of OHIP is the oral health interpretation of ICIDH via Locker’s model of oral health. According to this model, oral disease can have various consequences in a sequential manner, whereby impairments caused by the oral disease lead to functional limitations or discomfort/pain, which lead in turn to disability and ultimately to handicap (Figure 2) (Locker, 1988). In some cases, impairments and discomfort/pain can directly lead to handicap, and intervening variables such as personal and environmental factors can play an important role in oral health (Baker, 2007). The main merit of such a conceptual model is that it no longer defines
oral health as merely an absence of disease but also as a state that includes functional, social and psychological well-being, thereby enabling researchers to focus on a range of physical, psychological and social roles (Locker, 1988).

![Diagram of Locker's model of oral health]

In accordance with its theoretical framework, the aim of OHIP is to measure self-reports pertaining to impairment, disability and handicap due to oral conditions. In OHIP, 46 out of 49 items were factual statements identified from the personal interviews with 64 South Australian dental patients to reflect 7 domains hypothesized by Locker’s model of oral health. Three additional questions were added ad hoc by the researchers as none of the 64 participants were concerned with handicap, which Locker had identified as a domain of oral health (Slade, 1997).

The long-form OHIP-49 with 49 items was reduced to develop various shorter forms, two of which present 14 items each derived from two different methods: regression method and item-impact method. In the regression method, items are selected based on their prevalence and ability to predict overall scores (Slade, 1997) whereas the item-impact method used respondents’ ratings on the importance of each question (Locker & Allen, 2002). It is suggested that the regression method may be more suitable for studies whose aim is to discriminate...
between individual patients while the item-impact method might be better for describing the OHQoL of populations for epidemiological studies (Locker & Allen, 2002).

The OHIP-14 is scored and analyzed for individual theoretical domains or for the entire profile (Table 2) in one of two ways: (1) by simply counting items above the threshold response (e.g., “fairly often”), or (2) by adding up the Z scores of individual items (the individual score minus the mean of the population divided by the standard deviation of that population) (Slade, 1997). The former scoring method tends to underestimate scores because not many individuals report impact above the threshold, thereby violating assumptions of parametric distribution. Nevertheless, the summation scoring method is preferred to the latter Z-score method, which can be too laborious and complicated for practical purposes (Slade, 1997). Locker and Allen (2007) recommended that OHIP-14 be scored by the numbers of above-threshold responses out of the 14 items. Alternatively, the numeric values of Likert scale responses (e.g., 0 = never) for the 14 items can be combined to give total scores ranging from 0 to 56, higher values indicating more frequent oral health impact. However, the significance of overall profile scores has been questioned by MacEntee (2006) and again by Brondani and MacEntee (2007), as the same scores can be achieved by a different set of answers. For the same reason, the profile scores are frequently reported with domain scores rather than as a single total score.
Table 2. The original short form of the Oral Health Impact Profile (OHIP-14) questions and theoretical domains (Slade, 1997)

<table>
<thead>
<tr>
<th>Theoretical Domains</th>
<th>OHIP-14 Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional Limitation</td>
<td>Q1. Have you had trouble <strong>pronouncing any words</strong> because of problems with your teeth, mouth or dentures?</td>
</tr>
<tr>
<td></td>
<td>Q2. Have you felt that your <strong>sense of taste</strong> has worsened because of problems with your teeth, mouth or dentures?</td>
</tr>
<tr>
<td>Pain &amp; Discomfort</td>
<td>Q3. Have you had <strong>painful aching</strong> in your mouth?</td>
</tr>
<tr>
<td></td>
<td>Q4. Have you found it <strong>uncomfortable to eat any foods</strong> because of problems with your teeth, mouth or dentures?</td>
</tr>
<tr>
<td>Psychological Discomfort</td>
<td>Q5. Have you been <strong>self-conscious</strong> because of your teeth, mouth or dentures?</td>
</tr>
<tr>
<td></td>
<td>Q6. Have you <strong>felt tense</strong> because of problems with your teeth, mouth or dentures?</td>
</tr>
<tr>
<td>Physical Disability</td>
<td>7. Has your <strong>diet been unsatisfactory</strong> because of problems with your teeth, mouth or dentures?</td>
</tr>
<tr>
<td></td>
<td>Q8. Have you had to <strong>interrupt meals</strong> because of problems with your teeth, mouth or dentures?</td>
</tr>
<tr>
<td>Psychological Disability</td>
<td>Q9. Have you found it <strong>difficult to relax</strong> because of problems with your teeth, mouth or dentures?</td>
</tr>
<tr>
<td></td>
<td>Q10. Have you been <strong>a bit embarrassed</strong> because of problems with your teeth, mouth or dentures?</td>
</tr>
<tr>
<td>Social Disability</td>
<td>Q11. Have you been <strong>a bit irritable</strong> with other people because of problems with your teeth, mouth or dentures?</td>
</tr>
<tr>
<td></td>
<td>Q12. Have you had <strong>difficulty doing your usual jobs</strong> because of problems with your teeth, mouth or dentures?</td>
</tr>
<tr>
<td>Handicap</td>
<td>Q13. Have you felt that life in general was <strong>less satisfying</strong> because of problems with your teeth, mouth or dentures?</td>
</tr>
<tr>
<td></td>
<td>Q14. Have you been <strong>totally unable to function</strong> because of problems with your teeth, mouth or dentures?</td>
</tr>
</tbody>
</table>
2.1.2.2 Criticisms of Locker’s Model and OHIP

Although Locker’s model is one of the most widely accepted theoretical frameworks of oral health, it is not without limitations. Among the criticisms are its negative portrayal of health; the linear presentation of disease states; and the unequal power distribution between doctor and patient. Owing to its roots in Parson’s sick role theory, Locker’s model of oral health has negative classification labels, such as impairment, disability and handicap, and ignores the positive end of the health spectrum including the social benefits of positive oral aesthetics, for example. This seems to be a common problem among OHQoL measures that are based on the early ICIDH classifications. In review of 17 of the existing OHQoL scales, almost all were found to have items with negative connotations (Brondani & MacEntee, 2007). Consequently, few, if any, ICIDH-based scales capture the full range of oral health states and experiences, such as positive oral health behaviours and beliefs as well as coping and adapting to the oral problems (MacEntee, 2006; Brondani & MacEntee, 2007). Moreover, ICIDH-based scales have retained a vestige of the sick role theory, which dictates that it is the exclusive role of health professionals to reverse the progression of the disease. As a result, ICIDH does not allow for diseases that have unknown causes, nor does it consider convalescence, coping and adapting strategies or reciprocal influences among social, psychological, and biological processes. The negative connotations associated with disability and handicap might further encumber people with disabilities, causing loss of self-esteem and the development of defensive coping and adapting (Barnes & Mercer, 1996).

Additionally, the unequal patient-doctor relationship of ICIDH ignores lay perspectives on oral health. As noted by Berkonovic (1972) and Conrad (1990), the sick role theory offers a rather limited scope of health and illness from the standpoint of outsiders, such as health care
providers or medical sociologists, while overlooking the subjective perceptions and expectations of those who experience and manage their illness. This type of patient-doctor relationship is also reflected in the sociomedical definition of oral health: “the objective which the dental profession has in mind in assisting the individual to achieve a satisfactory state of function, comfort, and appearance” (Gerrie, 1947, p. 1318). In this reflection, a normal mouth is based on the opinion of a dentist, irrespective of the patient’s concerns (McCollum, 1943; MacEntee, 1997). Given these limitations, the current paradigm of oral health is gradually becoming a biopsychosocial conceptualization.

2.1.3 Biopsychosocial Model

In keeping with WHO’s definition of health as a state of complete physical, mental, and social well-being – and not merely the absence of disease (WHO, 1977), Engel (1977) offered an alternative route to expanding the traditional biomedical model. His overarching framework, better known as the biopsychosocial model, is a comprehensive multi-systems model in which a combination of biological (e.g., immune system), psychological (e.g., depression) and social factors (e.g., socioeconomic status) contribute to health in various contexts. The main advantage of the biopsychosocial model is its ability to analyze individual responses to the same disease within a multifaceted perspective on health. Further improvements in the biopsychosocial model included consideration of a patient’s subjective experience and active roles in the traditional patient-doctor relationship (Borrell-Carrio et al., 2004). The emergence of the biopsychosocial view of health demanded expansion of the original ICIDH framework (WHO, 1980) in order to incorporate quality-of-life (Jette, 1993; Tennant & McKenna, 1995) and the lay perspective of patients (Peters, 1996). The WHO (2001) later revised the ICIDH in the context of the International Classification of Functioning, Disability and Health (ICF) –
provisionally titled ICIDH-2 – to include new dimensions, such as environmental factors and participation in social activities. However, the new dimensions of health have yet to be reflected in the currently available Socio Dental Indicators (SDIs) (Brondani & MacEntee, 2007).

In accordance with the transition to the biopsychosocial framework in medicine, the concept of oral health acquired a more comprehensive definition: "a standard of the oral and related tissues which enables an individual to eat, speak and socialise without active disease, discomfort or embarrassment and which contributes to general well-being" (UK Department of Health, 1994, p. 55). Moving away from the sociomedical definition that focuses on social roles, the new biopsychosocial definition of oral health offers an integrated and holistic view that is intricately linked to health and quality of life (QoL) by referring to “an individual's perceptions in the context of their culture and value systems, and their personal goals, standards and concerns” (WHO, 2011). As presented earlier on, the need for quality-of-life measures sensitive enough to detect changes in oral health status and the psychosocial impacts of oral problems also led to the development of OHQoL scales, which have been identified also as Socio Dental Indicators (SDIs) (Cohen & Jago, 1976).

In summary, both the sociomedical and biopsychosocial models influenced the contemporary conceptualization of OHQoL. However, at the present time, most OHQoL scales – including the OHIP – are based on sociomedical frameworks, such as the older ICIDH, rather than the newer ICF. In my study, because of the theoretical roots of the OHIP, I used the former framework for the purpose of content-validating exercise in accordance with Locker’s ICIDH-based model.
2.1.4 OHQoL Measurement in Korean

Recently, many OHQoL measures have been made available in Korean (Bae et al., 2007; Jung et al., 2008; Lee et al., 2008; Shin & Jung, 2008; Ha et al., 2009; Kim et al., 2009), including the OHIP-14 (Slade & Spencer, 1994). There are at least three different variations of OHIP-14K (Bae et al., 2007; Kim & Min, 2008; Kim & Min, 2009). Even though they seem analogous (Appendix B.2, B.3, & B.4), it is confusing to have multiple variations of the same scale (van Windenfelt et al., 2005), and possibly in violation of the original OHIP-14 copyright (Hunt et al., 1991). I will use the translation of the OHIP-14 by Bae and colleagues (2007) as the official version of the Korean OHIP because it was the only version whose cultural adaptation and validation have been documented in the literature (Bae et al., 2007).

In order to appraise cross-cultural measurement of OHQoL in Korean, I will first describe the theories of validity and cultural equivalence by attempting to answer my first research question: how are validity and cultural equivalencies conceptualized in cross-cultural settings?
2.2 Validity Theory

Validity is defined as confidence that the scale measures what it purports to measure in the target culture or population (Sandelowski, 1986). Despite this simple definition, there have been many controversies surrounding the concept of validity; however, the Trinitarian and Unitarian theories dominate.

2.2.1 Trinitarian Concepts of Validity and OHIP-14K

Owing to its roots in educational and psychological testing, traditional validation efforts are centred on demonstrating the utility of scales by correlating its scores with external variables or criteria, such as reading skills (Thurstone, 1932). Thorndike (1931) and Jenkins (1946) criticized this validation approach on the grounds that the statistically driven approach does not capture the relevance of criteria to the measurement purpose or, more importantly, to the validity of the criteria themselves (Sireci, 1998). To overcome these limitations, Kelley proposed that criterion-related evidence be combined with professional judgment (1927), which led subsequently to the emergence of the Trinitarian concept of construct, criterion, and content validity whereby:

- Construct validity is the ultimate goal of psychometric measurement by indicating how accurately a scale reflects the construct of interest theorized by an underlying model selected by the scale developer (Hubley & Zumbo, 1996; Devellis, 2003). In my thesis, the construct validity of OHIP-14K indicates how well the construct of OHQoL hypothesized by Locker’s model of oral health is measured in a Korean context. Construct validity is typically asserted in one of three ways. The first way is through correlation with a gold standard, although construct validity of OHIP measurement in
English or in other languages has not been established because there is no accepted gold standard for OHQoL (McGrath & Bedi, 2001). The second way is by the statistics of factor analysis “whose common objective is to represent a set of variables in terms of a smaller number of hypothetical variables.” (Kim & Mueller, 1978, p.57) However, the accuracy of factor analysis is only as good as the validity of the data being modelled. Before performing a factor analysis, any potential bias in the content of the scale must be addressed to identify or explain the causality of score variations (Rahman et al., 2010). The third and final way is by establishing convergent validity by assessing the extent to which two or more scales of the same trait correlate (Larcker & Lessig, 1980; Hubley & Zumbo, 1996).

- Criterion validity is an empirical correlation between a test score and concurrent or predictive criteria (Messick, 1989), such as behaviours, personalities, abilities or outcomes, depending on the measurement objective of the scale. The current validation strategies for OHIP rely heavily on criterion-related variables, which vary greatly between the different language versions. In general, one tests for score differences between categories of respondents that are expected to differ (e.g., dentate vs. edentulous populations). However, criterion validity does not provide a theoretical explanation of the relationships between or causality of observed scores. With criterion-related evidence, it is not certain that what is being measured is indeed the construct of interest (e.g., OHQoL) or that the observed score is meaningful to the respondents (Sim & Waterfield, 1997).
Content validity refers to expert judgment about how the scale measures the situations or subject matter from which conclusions are drawn (Sireci, 1998). Unlike criterion validity, it is concerned with the clarity, relevance, and precision of the questions (Rulon, 1946; Mosier, 1947; Hubley & Zumbo, 1996; Yagmhmale, 2003). It serves as “an important indicator of the instrument’s quality” (Polit & Beck, 2006, p. 489) and as “a link between abstract theoretical construct and measurable indicators” (Wynd et al., 2003, p. 508). Therefore, content validity is crucial evidence of construct validity (Anastasi, 1988; Messick, 1989; Ding & Hershberger, 2002).

To date, the construct validity of OHIP14-K has been supported by Bae and colleagues (2007), who compared the observed OHIP scores to the criterion-related variables, such as perceived dental needs and number of natural teeth. Other methods for testing the content validity of OHIP-K have been suggested, including consultation with lay participants, or relying upon the judgment of subject-matter experts. A participant’s perception of the scale’s appropriateness is known as the “face validity” (Nevo, 1985). For instance, Montero-Martin and colleagues (2010) examined the face validity of the Short Form Health Survey (SF-36) to validate its content and found that its content domains were indeed relevant to target populations. However, face validity is extremely difficult to quantify and is insufficient as a psychometric property because judgments of validity are based mostly on the appearance of items rather than on their deeper theoretical foundation (Mosier, 1947; American Psychological Association, 1974). For this reason, face validity has been integrated with content validity (Haynes et al., 1995).

Another method of content-oriented validation involves the collection of expert judgments by administering evaluation surveys to those with adequate experience and knowledge of disease
on quality of life (Lynn, 1986; Sireci & Geisinger, 1995; Beitz & van Rijswijk, 1999; Wynd et al., 2003; Hubley & Palepu, 2007). However, it has not been used for validating an OHQoL scale like the OHIP-14. The main advantages of this method are simplicity, ease of use, cost-effectiveness, speed of application, and quantitative assessment of content validity index (CVI), all of which help researchers to make decisions about retaining, revising or eliminating items from a self-report scale (Schilling et al., 2007).

Up to this point, three broad attributes of validity have been reviewed. Although this conventional view of validity is useful for describing different aspects of this phenomenon, it might also benefit to discuss the unifying validity framework.

### 2.2.2 Unitarian Concept of Validity

Within the Unitarian framework, criterion-related, content-related and construct-related validity help to demonstrate construct validity (AERA, APA & NCME, 1999). In other words, construct validity of a measurement is supported by theoretical considerations and empirical evidence (Lawshe, 1975). There are two views of the type and amount of evidence required to validate a psychosocial measurement. Kane (2009) argued that validity should depend on an assessment of the purpose of the measurement rather than on the scale itself. The predictive attributes of the measurements, if they exist, are what really matter.

Zumbo (2009) contended that criterion-driven validation overlooks the importance of theoretical explanations for the scores and for item response variations (2009). Vickery (1998) took this argument a step further by claiming that construct validity and reliability are meaningless without content-related evidences because content validity appeals to reason and theoretical justification. This view is shared by Messick (1989, p. 13), who defined validity as “an
"integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the interpretations of assessment scores". In this context, content validity, the principle for theoretical explanation of the measurement, is in fact an attribute of construct validity.

As noted by many validity theorists including Sireci (1998) and Schouwstra (2000), this Unitarian framework is gaining ground as a suitable theoretical basis for studying validity because construct validity cannot be established if the content of the scale is not valid. Another main tenet of this framework is that construct validity can be strengthened by minimizing threats, including *construct underrepresentation* (i.e., a scale fails to capture important dimensions of the construct) and *construct irrelevant variance* (i.e., variance due to non-related constructs or method variance) (Messick, 1989).

In my thesis, I adopted Messick’s Unitarian theory of validity as the conceptual framework in which content- and criterion-related validities are integrated as a crucial piece of evidence for the measurement’s construct validity (Messick, 1989; Yaghmale, 2003). An investigation of the content of the OHIP-14K, however, requires further discussion of another closely related but distinct concept: cross-cultural equivalence theory.

### 2.3 Cross-Cultural Equivalence Theory

A cross-cultural comparison of OHQoL provides cultural perspectives on the conceptualization and measurement of oral health. Central to the comparison is the notion of cultural equivalency. Because the accuracy of self-reported data is incumbent on the quality of cultural adaptation, linguistically and culturally equivalent scales are needed in multi-regional research. However, establishing cultural equivalence between two different-linguistic versions of the same scale is often challenged by the conflict between fidelity and cultural appropriateness (Corless et al.,
There are two competing views on the extent of accommodating cultural appropriateness in the translation in a way that challenges its fidelity to the original version: the *emic* and *etic* perspectives. The emic perspective interprets phenomena within a culturally specific context that is significant to its members, whereas the etic perspective studies cultural phenomena from an outsider’s point of view (Pike, 1967; Berry, 1990). The emic perspective has the advantage of providing an insider’s view of culture by dealing with culturally relevant subjects, and preventing an imposition of the researcher’s values and biases. Hence, cultural adaptation of OHIP should consider the surface structure, such as ethnic/cultural characteristics, experiences, norms, values, behavioural patterns and beliefs of a target population, as well as the “relevant historical, environmental, and social forces” influencing the culture (Resnikow et al., 2000, p. 272). However, it is not yet known whether it is possible to penetrate native experiences to provide a compatible framework for cross-cultural comparisons (Cohen, 1974; Alegria et al., 2004).

The etic or culturally comparative approach observes through generalizations of cultural characteristics and values from an outsider’s perspective (Gzagnon & Tuck, 2004). It has been criticized for assuming an idealistic or universal application of the observer’s values and biases to all cultures (Cohen, 1974) and for encouraging cultural homogeneity by imposing ethnocentric constructs on other cultures (Alegria et al., 2004). Consequently it could lead to a culturally insensitive scale that lacks content and construct validity for that target population (Rogler, 1999). Detailed examples of each case will be presented in section 2.4.3, Construct Equivalence, on page 28.

Given the merits of etic and emic perspectives in cross-cultural studies, my investigation of OHIP-14K includes both. Such a dual perspective on cultural equivalence has been suggested
by many researchers (Cohen, 1974; Bravo et al., 1991; Canino & Bravo, 1994; Herdman et al., 1998; Jones et al., 2001), and it defines cultural equivalence as multi-dimensional properties of cross-cultural comparison conceptualized broadly in terms of method, item, and construct equivalence (Hunt et al., 1991; van de Vijver & Tanzer, 1997).

2.3.1 Method Equivalence

Method equivalence, also known as technical (Canino & Bravo, 1994) or operational (Herdman et al., 1998) equivalence, refers to the use of unbiased methods for designing and administering multilingual versions of the same measure (van de Vijver & Tanzer, 1997). Method equivalence is one of the most overlooked aspects of cultural adaptation (Sireci et al., 2006), and to date, there is no benchmark for standardizing the format of the scale or the instructions in different languages. Potential threats to a scale’s method equivalence include biases related to the scale or administrative methods.

2.3.1.1 Scale Bias

Scale bias refers to cross-cultural differences in scores due to differences in the format of a questionnaire especially relating to the way in which possible responses are presented (van de Vijver & Tanzer, 1997). Consequently, the structures and format of the original and the translated versions must be preserved while accommodating the cultural needs of the target participants. For example, participants who are unfamiliar with the format of a Likert scale might find it confusing to use (Bracken & Barona, 1991). A scale bias could also occur during the cultural adaptation of a questionnaire if Likert categories are narrow in scope or ambiguous. For example, the Likert scale of the original English version of OHIP allows respondents to answer each question with five choices plus an optional “don’t know” (DK) response (Slade &
Spencer, 1994), yet the OHIP-K does not offer the DK option (Bae et al., 2007). The DK options seem to eliminate guessing which should improve the accuracy of the responses. In particular, answering the OHIP questions is a memory-dependent task as it requires respondents to recall related events from the past 3, 6 or 12 months, and the DK options is useful for respondents who cannot recall the events or simply don’t know the answer to the question (Ritter & Sue, 2007). In support of this idea, Courtenay and Weidemann (1985) found that DK option eliminates guessing and produces more accurate answers on Palmore’s Facts on Aging Quiz. On the other hand, there are concerns that DK responses is difficult to interpret for researchers (Holbrook, 2008) and that it might distract respondents from making a choice from the more definite possibilities (Krosnick, 1991; Gilljam & Granberg, 1993). However, these findings mostly apply to opinion-type questions about personal dispositions, attitudes, beliefs or agreements rather than to factual questions such as those posed by the OHIP. Indeed, the low non-response rates from the sampled populations in many oral health surveys suggest that the DK options in the OHIP are less likely to be abused by respondents (Locker, 1993). Overall, it seems that keeping the response formats uniform across different-language versions is necessary to minimize the scale bias for valid cross-cultural comparisons.

Another example of a scale bias can be found in the Likert scales used for another existing version of the OHIP-14K proposed by Kim and Min (2008). In this Korean translation of “very often, fairly often, occasionally, hardly ever and never” in the original English OHIP, we see that the choices “hardly ever” and “never “ became “it’s not like that” and “it is never like that” (Appendix B.4). Not only do these response categories have unequal intervals, but they also offer little or no information about the frequency of events.
2.3.1.2 Administration Bias

An administration bias is caused by different instructions or other communications for using the instrument (Feinstein, 1987; van de Vijver & Tanzer, 1997). Yet the instructions given in the OHIP-14K differ significantly from their English counterparts in recall periods, frequency of oral problems, and range of oral health problems (Bae et al., 2007).

The one-year recall period of oral health-related events in the original OHIP was reduced to three months in the OHIP-14K. Shorter recall periods tend to yield more accurate responses but underestimate the prevalence of problems, and vice versa for longer recall periods (WHO, 2003). Hence, longer recall periods are ideal for describing chronic oral problems, but may miss acute changes in oral health and impose a cognitive burden on respondents, especially elderly respondents with some cognitive impairment. This may support the shortened recall period for the OHIP-14K; however, the difference in recall periods might bias cross-cultural comparisons. For instance, the lower test-retest reliability (intra-class correlation coefficients) for the OHIP-14K (0.40–0.64) compared with German (0.63–0.92) and Chinese versions (0.63–0.92) lead Bae and colleagues to conclude that OHIP-14K might be measuring acute changes in respondents’ oral health status rather than their oral health status overall.

Unlike the original version of OHIP-14, the OHIP-14K gives a defined set of frequencies per week for each Likert response (e.g., very often = more than 3 times a week). The additional instructions could impose a cognitive burden on respondents and could also reduce bias as it eliminates the variation caused by individual perceptions and judgments of the frequency of the events. Nonetheless, the objective criteria can have unintended consequences by introducing researchers’ biases into their measurement of OHQoL when the aim is to reflect the
participants’ subjective perceptions on the impact oral health rather than on their oral health history. There is a chance that response categories arbitrarily set by researchers may be different from those determined by the respondents. For example, “more than 3 times a week” could imply occasional happenings for some, and high frequency for others. Thus, the inclusion of specific frequencies in one language and not in others could lead to misleading or at least incomparable responses across cultures.

Another source of method bias is the limited scope of oral problems defined by OHIP-14K. It only asks about problems with teeth and gums whilst the OHIP includes problems associated with the tongue, dentures or mouth in general. This instruction bias could underestimate the oral problems of Korean respondents. Also, the way in which the directions are presented to respondents may affect their responses. In the original OHIP, many of the oral health-specific questions overlap with systemic health or psychological conditions (e.g., trouble pronouncing words due to physical impairment of vocal cords or muscle tone, not due to oral problems). Therefore, every OHIP question in English has the phrase “because of problems with your teeth, mouth or dentures” to define the topic of interest, while this phrase was omitted from the OHIP-K items. Repeating the instructions in every question may be necessary to improve the respondents’ comprehension and responses (Smith, 2004; Somekh & Lewin, 2005; Dykema et al., 2010).

In light of these findings, method biases, therefore, can have pervasive effects on various elements of OHQoL measurement and adversely affect the outcomes of cross-cultural comparisons. But because method biases do not affect correlations, method bias is undetected by most statistical analyses and is almost never addressed in cultural adaptation or validation.
studies (van de Vijiver & Poortinga, 2006). While it remains unknown whether the translated versions of the OHIP have standardized measurement methods, the authors have suggested that collective expert judgment could be used to detect potential biases in the instructions, response format, and other features of the scale.

### 2.3.2 Item Equivalence

The quality of self-reports depends on how clearly and accurately the questions are posed, so language and culture must be considered carefully when translating a scale for different cultural groups (Ellis, 1989). Even slight differences in wording can change meaning and elicit different responses (McTavish, 1997). The various forms of item equivalence between source and target versions of the same scale have been consolidated into five categories: semantic equivalence, experiential equivalence, functional equivalence, scalar equivalence and conceptual equivalence (Guillemin et al., 1993; Hui, 2008).

1. **Semantic equivalence** means the similarity in linguistic meaning across cultures, and it can be attained by ensuring faithfulness to the original texts and cultural appropriateness for a target population through multiple forward-back translations and committee reviews (Canino & Bravo, 1994; Herdman et al., 1998).

2. **Experiential equivalence** refers to the extent to which experienced situations described by target items are equivalent to those of the source items (Canino & Bravo, 1994).

3. **Functional equivalence** suggests that target texts convey the meaning of the source texts in a way that preserves the purpose and functions of the original texts (Bullinger et al., 1993).

4. **Conceptual equivalence** implies that identical concepts have the same meaning across cultures (Usunier, 1998) and is achieved when questionnaires have the same relationship
with the underlying concept or theoretical dimensions of both cultures (Herdman et al., 1998).

5. **Scalar equivalence** means that the scale has same measurement unit across cultures (Canino & Bravo, 1994; Lee et al., 2009). As noted by Bartram (2011), scalar equivalence is difficult to achieve because score differences between cultural groups could be due to the true score variation as well as systematic scale-related biases. Therefore, unlike other types of item equivalence, scalar biases can be tested only by administering the scale and interviewing respondents in order to verify the same degree of construct indeed produce equivalent scores. Although scalar equivalence was beyond the scope of my content-validation, its implications in relation to my findings are examined in the Discussion.

In the literature, various solutions have been proposed to minimize the multiple types of item biases presented above, including the adoption of items without modifications, minor modifications to items, replacement of items, and elimination of items deemed inappropriate. Some problematic items would require re-translation or further verification from bicultural experts (Herdman et al., 1998).

2.3.3 **Construct Equivalence**

The ultimate goal of cross-cultural adaptation is to establish construct equivalence, the degree to which the same construct is measured across populations of interest (van de Vijver & Tanzer, 1997; Herdman et al, 1998; Kristjansson et al., 2003). Potential threats to construct equivalence include biased sampling of behaviours relevant to the construct, incomplete representation of the construct, and a lack of construct overlap (Matsumoto & van de Vijver, 2011). For instance, the
biopsychosocial consequences of oral health can vary greatly among cultures depending on the significance given to oral health. It is also suggested that cultural climate can control the expression of social disability and the perception and tolerance of pain (Nayak et al., 2000). Allison and colleagues (1999) found substantial conceptual differences between Australian and Canadian judges in their perceptions of item severity. In more extreme cases, the construct of interest might not be comparable or directly transferable across cultures (Colress et al., 2001). For instance, despite their shared history and language, North and South Koreans have different cultures due to the partition of Korea, and the theoretical foundations underlying OHIP-14K may not respond the same way in the two countries. Therefore, apart from the linguistic factors, construct equivalence should be studied in a broader cultural context, taking into account cultural or sub-cultural factors such as social background, education, ethnicity, age, and gender. Duarte and Rossier (2008) suggested that the cultural approach to construct equivalence involves ensuring that the construct is both relevant to and observable in the target group before adjusting scales.

In sum, Figure 3 illustrates various aspects of validity and equivalency for cross-cultural measurement and comparison of OHQoL. As can be seen from the diagram, I found that construct equivalence of the measurement is supported by evidence of both validity and equivalence. To find out what evidence there is to support construct equivalence of OHQoL measurement, I compared and contrasted a variety of methods used to adapt and validate the OHIP-14 in the next chapter.
Figure 3. Various aspects of validity and equivalency for cross-cultural measurement and comparison of OHQoL
Chapter 3: Systematic Synthesis of the Literature

3.1 Search Strategy

1) The systematic synthesis I used is considered to be primary research with a clearly defined research question: To what extent does the Korean version of the short-form Oral Health Impact Profile (OHIP-14K) exhibit content validity and cultural equivalency in Korean? Papers were selected based on their relevance to cultural adaptation and validation of OHIP. As the original English version of OHIP was developed in 1994 (Slade & Spencer, 1994), a literature search was performed to identify articles published from 1994 onward through Medline, EMBASE, and CINAHL, using the keywords “OHIP or Oral Health Impact Profile,” “cultural adaptation”, and “validation” in subject headings, titles, and abstracts. The examples of search coding schemes are presented in Appendix A.1.

I included publications showing: (1) translation and cultural adaptation from the source English version of OHIP to various languages; (2) all types of validity evidence for OHIP obtained in different cultures; and (3) cross-cultural comparisons of responses using different language versions of OHIP. I excluded articles published in languages other than English and Korean along with others that did not describe the process of how the English version was adapted to the target culture.

I read the titles and abstracts of the papers identified in this first step of the search for relevance to the second research question, but only the twenty two studies that were considered relevant to cross-cultural measurement of the OHIP were read fully and analyzed critically (Figure 4).
Phase 1: Abstract Screening

Phase 2: Full Article Inspection

The search identified 19 papers describing various cultural adaptation and validation procedures for the OHIP. The most widely used methods were found to be forward and back-translations, pilot testing and post-hoc statistical analysis, usually following the pattern presented in Figure 5.
Source English OHIP

Forward Translation

Back-Translation

Committee Review

Pilot Testing

Target Language OHIP

Figure 5. Flow chart of cultural adaptation stages for OHIP
In the following sections, I will explain each cultural adaptation method in detail with its relative strengths and weaknesses.

3.1.1  Forward Translation

With a notable exception of the English version of OHIP-14 developed specifically for Scottish populations (Fernandes, et al., 2005), 18 other versions were forward-translated from the source. In conjunction with the forward translations of the original English version, the German (John et al., 2002) and Japanese (Yamazaki et al., 2007) versions were developed de novo from personal statements obtained from the respective populations to establish content validity. Forward translation, a process of translating from one language to another, is usually performed by one or more bilingual health professionals familiar with the content and goals of the study to ensure that the clinical meaning is appropriate (King et al., 2011; WHO, 2011).

There are two main translational approaches in health research: literal and conceptual (WHO, 2011). Literal translations follow the original texts as closely as possible in tone, grammatical structure and idioms, but are not always semantically equivalent or in context. On the other hand, conceptual translations can be too liberal or lenient, which can weaken their precision and fidelity to the original texts. Therefore, forward translations alone do not necessarily establish equivalence between the source and target languages (Maneesriwongul & Dixon, 2004). Nor is forward translation a bias-free process because there are many words and phrases in one language that have no equivalent meaning in other languages, and their interpretation is at the discretion of the translator. As a result, a translation bias can arise from a number of factors including: bilingual proficiency and experience (Wild et al., 2005; Hambleton, 2006); a lack of training in the construction of psychometric measures (Hambleton & Patsula, 1998); and an
inadequate knowledge of OHQoL constructs and their underlying theories. For these reasons, many studies adopted a precautionary step to avoid translational errors by using multiple and independent translators. Examples of this approach include the multiple forward-translations used to create the Italian, Spanish, Dutch, and Brazilian versions of the OHIP by two or more translators working independently and using a mediator or an expert to resolve disagreements (Segu et al., 2005; Lopez & Baelum, 2006; Pires et al., 2006; Meulen et al, 2008). In contrast to single forward translations, multiple forward translations reduce the risk of biased translation caused by one person’s writing style, speech habit or misinterpretations (Wild et al., 2005).

3.1.2 Back-Translation

The most frequently used judgmental procedure for detecting bias in forward translations is back-translation, whereby another bilingual translator independently translates the forward translation back into the source language (Guillemin et al., 1993). The back-translations are then compared to the original version for semantic and conceptual equivalence (Brislin, 1970; Hambleton, 2006). However, there is debate, over who should perform the back-translations. Van Winderfelt and colleagues (2005) recommend translators with knowledge about the subject of interest. In contrast, Guillemin (1995) argues that back-translations should be performed by translators who have no familiarity with the subject. Hence, it might be necessary to employ both types of back-translators.

Even close similarities in language between the back-translation and the original does not necessarily produce cross-cultural equivalence because translators may share the same set of rules for translating non-equivalent words or compensate for poor forward translation by using a literal translation with no consideration for conceptual differences between the texts (Brislin, 1970; Stansfield, 2003). Therefore, to make a valid inference about the equivalence between the
source and target versions it is also necessary to know which type of back-translation was performed. Like forward translation, back-translations can be performed literally or conceptually. The former attempts to produce semantic equivalence while the latter aims to reflect the same construct in different cultures (Herdman et al., 1997). Wild and colleagues (2005) suggested that literal back-translations can be more suitable for medical surveys that are relatively objective as opposed to the more subjective QoL-related items, which require conceptual back-translations. However, unlike most QoL scales, most of the OHIP questions ask about oral health-related experiences (e.g., difficulty chewing), so a conceptual back-translation may not be sensitive enough to detect a potential misrepresentation of the original concepts in forward translations. For this reason, Sen and Mari (1986) recommend both literal and conceptual back-translations to ensure linguistic and conceptual equivalence between the source and target versions.

3.1.3 Committee Translation

In contrast to a single forward translation, multiple translations can be performed by a committee of independent translators to detect an individual translator’s errors (Brislin, 1970; Beaton et al., 2002). Parallel and serial translations of the OHIP have been assessed in this way. In the parallel method, independent translators produce forward and back-translations in a parallel fashion, which are later reconciled into one final version by another translator (Brislin, 1970). Twelve out of 19 versions of OHIP were developed using this parallel method with at least two forward translations and two back-translations.

The serial or multiple pass method uses a group of bilingual translators who take turns in increasing order of expertise to review each other’s translations and make necessary amendments (Gouadec, 2007). The authors of OHIP-14K opted for this method with a panel of
four Korean dentists who reviewed the forward translations in Korean. However, committees can be costly and time-consuming to form with three or more bilingual translators, and they do not necessarily operate without a shared bias or eliminate the potentially distorting influence of a hierarchy of power among members (Drennan et al., 1991; Maneesriwongul & Dixon, 2004).

**3.1.4 Pilot Testing**

Pilot testing a provisional version of OHIP with a small subset of the monolingual participants give the participants an opportunity to judge and if necessary modify the scale. For instance, OHIP-14K was pilot tested with a convenience sample of 10 Korean adults, and, according to the authors, no comprehension difficulties were reported on the OHIP-14K questions. However, this will not necessarily establish the content validity or cross-cultural equivalency of a scale because of the limited advice expected from monolinguals.

**3.1.5 Statistical Validation**

In general, post-hoc statistical analyses involve calculation of internal consistency and test-retest reliability. Internal consistency with Cronbach’s alpha (α) statistics evaluates the interrelatedness of items to the construct of interest, while test-retest reliability refers to the stability of scores obtained at two different times for the same population (Hubley & Zumbo, 1996). However, it is not always indicative of the quality of the scale because an extremely high alpha value (α > .90) can also indicate redundancy of items (Streiner, 2003) and low validity as a result of the scale being too narrow and specific (Kline, 1979). Streiner (2003) recommends that the maximum Cronbach’s alpha value not exceed 0.9. By the same measure, a very high alpha value coupled with high inter-item correlation of the Dutch version of OHIP has led to the conclusion that the five out of seven domains were redundant (Meulen et al, 2008).
Test-retest reliability refers to the consistency of psychometric measurements; however, it does not address the problems inherent to the scale because re-administering the same scale only reintroduces the same bias. Also, if the test-retest intervals are too far apart, it might measure changes in subjects’ physical state, severity perception or point of reference, rather than a real change in QoL. Changes in internal standards, values or conceptualizations, coping strategies, social comparison, social support, prioritization, expectations and spiritual practice can occur over time to confuse the results of the test-retest (Sprangers & Schwartz, 1999).

These steps of cultural adaptation and validation used for the 19 language versions can be summarized in Table 3.
Table 3. The various methods used to translate and culturally adapt the OHIP

<table>
<thead>
<tr>
<th>Authors</th>
<th>Language</th>
<th>Version</th>
<th>Forward &amp; Back-Translation</th>
<th>Pilot Study</th>
<th>Committee/ Multiple Translation</th>
<th>De novo Development¹</th>
<th>Long Form²</th>
<th>Short Form³</th>
<th>Content Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wong et al., 2002</td>
<td>Chinese</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Meulen et al., 2008</td>
<td>Dutch</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Allison et al., 1999</td>
<td>Canadian</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>John et al., 2002</td>
<td>German</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Papagiannopoulou et al., 2012</td>
<td>Greek</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>Kushnir et al., 2004</td>
<td>Hebrew</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Szentpetery et al., 2006</td>
<td>Hungarian</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Segu et al., 2005</td>
<td>Italian</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Yamazaki et al., 2007</td>
<td>Japanese</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Bae et al., 2007</td>
<td>Korean</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>Kenig &amp; Nikolovska, 2012</td>
<td>Macedonian</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>Saub et al., 2005</td>
<td>Malay</td>
<td>O</td>
<td>O</td>
<td>N/A</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Pires et al., 2006</td>
<td>Portuguese</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Fernandes et al., 2005</td>
<td>Scottish</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>Stancic et al., 2009</td>
<td>Serbian</td>
<td>O</td>
<td>O</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>Ekanayake &amp; Perera, 2004</td>
<td>Sinhalese</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>Lopez &amp; Baelum, 2006</td>
<td>Spanish</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Montero-Martin et al., 2009</td>
<td>Spanish</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>Larsson et al., 2004</td>
<td>Swedish</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

¹ De novo method refers to the new development of a questionnaire according to the same method used for the original English version of OHIP (John et al., 2002).

² The long-form OHIP, which was its original length, has 49 items (Slade & Spencer, 1994).

³ The short-form OHIP refers to OHIP-14, which has 14 items (Slade, 1997).
It is evident from the systematic synthesis findings that the current cultural adaptation and validation strategies do not fully address the content validity and cultural equivalency of the OHIP-14.

3.2 Content Validity

First, cross-cultural measurement of OHQoL must demonstrate content validity for the target population. As described earlier, content validity means the relevance and representativeness of scale content (e.g., items or questions) to the construct of interest and its appropriateness for the target population (Beck & Gable, 2001). Whilst a variety of sources of validity evidence is needed for scale validation (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999), many OHQoL studies have neglected efforts to examine the theoretical aspect of cross-cultural measures on the assumption that the known psychometric properties of an originally validated scale, such as content and construct validity, will remain constant across cultures through cultural adaptation.

Nevertheless, content validity is not a fixed property of a scale and requires ongoing evaluation within the context of multiple sources of evidence (Sireci, 1998, 2007). In some cases, the judgments on content validity made by the originators are not generalizable to cultures outside the mainstream Western culture in which they were made (Rogler, 1999). For instance, linguistic equivalence may not guarantee content validity for another culture. To illustrate this point, consider the following case: when Russia’s Supervisory Natural Resources team toured Korea to see mountain animals, a Korean official told the Russian representative that “Korea is very interested in Siberian tigers” (The Chosunilbo, 2009). The translation of this statement was misunderstood by the Russian delegates as a request for a tiger, so the zoo in Seoul received
three Siberian tigers. In this example, the literal translation from Korean to Russian did not make it clear that the Korean were simply admiring the animals and not asking that they receive tigers as a gift. In this light, content validity of a scale is only relevant to the intended culture (Hayness et al., 1995); therefore, linguistic and cultural modifications might be necessary to preserve the scale’s content validity (Debb, 2007).

The content validity of the OHIP-14 is especially vulnerable due to the small number of items measured (Locker & Allen, 2002), which can adversely affect the construct validity of its measurements (Haynes et al., 1995; Awad et al., 2008). Thus, further examination is necessary to fully evaluate the content validity of OHIP-14K.

3.3 Cross-cultural Equivalence

In order for OHIP to be a viable cross-cultural scale, it must demonstrate not only validity for the Korean population but also cultural equivalence with the original English version. The latter means semantic, colloquial, experiential, functional, and conceptual equivalence between the target and source versions (Guillemin, 1995) for unbiased measurement of the construct between different cultural groups (Cella et al., 1996). The current practices of cultural adaptation and validation are criticized for relying heavily on back-translation, reviews by lay panels, and on statistical techniques, whilst ignoring the conceptual equivalence of the scales (Herdman et al., 1998; King et al., 2011). Hence, the existing methods do not eliminate or control the cross-cultural biases that can adversely affect OHQoL measurements. The consequences can have far-reaching impacts on construct validity and equivalence. Addressing the comparability or cultural equivalence of a scale has been a great challenge for comparative health researchers (Liang, 2001; Ramirez et al., 2005) because no single approach is infallible to
the cross-cultural biases that pose threats to measurement validity in self-reported data (Hambleton & Patsula, 1998).

The method proposed by Allison and colleagues compares the relative unpleasantness of each impact judged by different cultural groups (1999). The authors found the consistency in the severity (or oral impact) rankings of the OHIP items by Australian and Anglophone- and Francophone-Canadian judges and concluded that OHIP is a cross-culturally valid tool. For instance, difficulty chewing was judged to be more unpleasant than the loss of sense of taste across the three cultural groups. However, the comparison of item impact alone may not yield adequate information on the linguistic or conceptual equivalence of items across cultures for cross-cultural comparison of OHQoL.

In fact, linguistic and cultural biases have previously been reported in various versions of the OHIP. When the question, “have you been self-conscious because of problems with your teeth, mouth or dentures?” was translated into Macedonian, the term self-consciousness had a very different meaning than in English, and half of the respondents could not answer the question (Kenig & Nikolovska, 2012). A similar problem with the meaning of self-consciousness was encountered in the Italian version of the OHIP (Segu et al., 2005).

Another method used for achieving conceptual and linguistic equivalence between two different-language versions of a scale is to administer both versions to a representative sample of bilingual respondents. For instance, strong positive correlations of scores obtained from Arabic, Spanish (Hunt, 1986), French (Hunt et al, 1991), and English versions of the Nottingham Health Profile were seen as evidence of equivalence. In a different study, positive Pearson’s correlations in the total scores of the Greek and English versions of the General Health Questionnaire have been used to support the conclusion that the Greek translation was
accurate and the two versions were equivalent (Garyfallos et al., 1991). However, when discrepancies in the scores arose – as in the Chinese version of the Menstrual Distress Questionnaire – the traditional comparative bilingual approach did not offer effective ways to improve the poorly adapted items or mediate disagreement between experts (Chang et al., 1999), and the causes of score differences is unknown. In another study by Son and colleagues (2000), significantly higher means were noted for two items on the Korean version of the Caregiving Satisfaction Scale when compared to the original English version: “caring for my elder helps keep her/him from getting sicker than she/he otherwise would” and “caring for my elder has taught me to distinguish the important things in life from the not so important”. The authors faced similar challenges when it came to explaining the score differences and modifying the items to improve cross-cultural equivalence.

There is a need for a structured method to first evaluate the psychometric properties of a translated scale and then to detect and derive solutions to potential biases for meaningful cross-cultural measurement and comparison. The health and educational research literature suggests that content-oriented validation has been used to establish content equivalence of alternate forms of various tests (Ding & Hershberger, 2002; Martin et al., 2010) and between multi-lingual versions of various scales, including the Quebec User Evaluation of Satisfaction with Assistive Technology Questionnaire (Brandt, 2005), Quality of Life Questionnaire (Cestari et al., 2006), Safety Attitudes Questionnaire (Devriendt et al., 2012), Diabetes Empowerment Scale (Shiu et al., 2003), Tobacco Acquisition Questionnaire and Decisional Balance Scale (Chen et al., 2003), Help Questionnaire (Kochinda, 2011), and Adolescent Teasing Scale (Liu, 2011).

Although it has never been attempted with an OHQoL scale, many cross-cultural researchers have suggested the possibility of submitting both source and target versions for evaluation by
bilingual subject matter experts (SMEs) to revise or remove items that are found to be non-equivalent (Geisinger, 1994; Chang et al., 1999; Beaton et al., 2002; Sireci et al., 2006). For instance, Kleiner and colleagues (2009) employed Chinese, Spanish, and French bilingual survey researchers to evaluate on a 7-point Likert scale the translation quality of the respective language versions of the National Survey of Children with Special Care Needs. The authors noted that a thorough assessment of the translation quality was necessary to enhance the naturalness of expression and ensure that the message is recreated with respect to modes of behaviour that are meaningful within the context of the target culture.

In summary, I can conclude from this literature review that a different validation approach is needed for assessing content validity and cross-cultural equivalency of OHIP-14K. I now present the empirical design I used to answer my third research question: *To what extent does OHIP-14K exhibit content validity and cultural equivalency?*
Chapter 4: Second Component – Empirical Exercise Methods

To assess the content validity and cultural equivalency of the OHIP-14K, I performed a content validation study based on quantitative judgments of bilingual SMEs who are familiar with both Korean and North American cultures. The significance of establishing the content validity and cultural equivalency of the OHIP-14K centres on certifying that (1) the instruction and items can be accurately understood by Korean respondents in their native language; (2) the OHIP-14K items are relevant to the hypothesized domains of OHQoL as explained by Locker’s model of oral health; and (3) the OHIP-14K is linguistically and culturally equivalent to the original English version.

In order to measure these properties of OHIP-14K, I adopted CVIs techniques suggested by Lynn (1985). Content validity assessment is generally divided into scale development and quantification phases. In my study, the development stage was omitted because OHIP-14K was adapted directly from the English OHIP-14, obviating the need for domain identification, item generation, and scale formation. I obtained directly from Bae and colleagues the version of OHIP-14K used for their validation study (2007). I carried out the quantification phase closely following the content validation procedures described by Crocker and Algina (1986):

1. Literature review of the psychological domains of OHQoL and its structured framework (as I already have presented throughout the first component of my dissertation, above);
2. Content validation exercise via selection of qualified SMEs;
3. Administration of CV questionnaires;
4. The data collection and evaluation of CVI for the OHIP-14K.
The last three steps are discussed as part of the empirical component of my dissertation. The following series of methods I employed are about the development, administration and analysis of Content Validity Questionnaires. First, however, I will present the selection process for the SMEs.

4.1 Selection of Bilingual Subject Matter Experts

The selection of experts should be based on careful consideration of their level of expertise and knowledge in the focal area of the research (Dubois & Dubois, 2000). Accordingly, I selected SMEs who:

(a) Had a dental licence for practice in Canada and practised general dentistry at least three times a week for a minimum of 10 years;
(b) Knew about oral problems and their impact on patients’ lives through relevant clinical experience;
(c) Understood both the Korean and the main-stream English population of Greater Vancouver Area;
(d) Used both English and Korean in their dental practice, but whose native language was Korean and who had resided mostly in Canada for the past 5 years;
(e) Were willing to participate in this study.

The authors of the OHIP-14K and anyone who had previous exposure to the OHIP were excluded from this study to maintain neutrality and to minimize potential confirmatory bias. The selection of recruited SMEs usually ranges from 3 to 10 experts for studies of content validation (Lynn, 1986; Sireci, 1998; Yaghmale, 2003; Hubley & Palepu, 2007; Domingues et al., 2011), but it also considers personal factors including expertise, scope of practice, and other practical considerations, such as the availability of qualified experts, the range of
expertise represented, and completion or drop-out rates (Hayness et al., 1995). Although there is no optimal number of SMEs for a content validation study, Gable (1986) suggested that a minimum of 10 judges was required to yield acceptably consistent responses and to avoid chance agreement. Lynn noted that the maximum number is unlikely to exceed 10 experts (1986).

Considering the brevity of OHIP-14K and the limited number of eligible SMEs to whom I had access, I aimed to recruit 10 SMEs. A convenience sample of 20 eligible SMEs was identified through the 2009 College of Dental Surgeons of British Columbia directory by seeking dentists with Korean names. Initial contact was made by phone or email followed by a personal visit to their clinical practices. Of the 20 SMEs identified in Greater Vancouver Area, 17 were contacted, as the other 3 could not be reached at their listed addresses. All potential SMEs received personalized cover letters detailing the research aim and rationale as well as the theoretical basis of Locker’s model (Appendix D.2). From these 17 potential recruits, 11 volunteered to participate while the others refused to participate because of their busy schedule. One of the selected SMEs eventually withdrew from the study because she felt that her Korean was not proficient enough for the content validation exercises. Therefore, I had 10 dentists on the SME panel, with one of them working in academia as well (Table 4).
<table>
<thead>
<tr>
<th>SME</th>
<th>Gender</th>
<th>Place(s) of Graduation</th>
<th>Clinical Experience (yrs)</th>
<th>Education Background</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>University of British Columbia; University of Temple; University of California, San Francisco</td>
<td>15</td>
<td>DDS, PhD, AEGD^a</td>
<td>Burnaby, BC</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>Seoul National University; University of British Columbia</td>
<td>26</td>
<td>DDS/MD</td>
<td>Burnaby, BC</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>Undisclosed</td>
<td>30</td>
<td>DMD</td>
<td>Coquitlam, BC</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>Yonsei University; University of Manitoba; University of British Columbia</td>
<td>19</td>
<td>DMD, MSc, PhD</td>
<td>Burnaby, BC</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>Southwestern University</td>
<td>14</td>
<td>DMD</td>
<td>North Vancouver, BC</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>University of London; Seoul National University; Korea University</td>
<td>41</td>
<td>BDS, DMD, MSc, PhD</td>
<td>Vancouver, BC/ Seoul, Korea</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>University of British Columbia</td>
<td>10</td>
<td>BDS, DDS</td>
<td>Burnaby, BC</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>University of Manitoba</td>
<td>15</td>
<td>DDS</td>
<td>North Vancouver, BC/ Lancaster, CA Coquitlam, BC</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>University of Pennsylvania; Columbia University</td>
<td>12</td>
<td>DDS, MA</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>M</td>
<td>Korea University; University of British Columbia</td>
<td>26</td>
<td>DDS (SNU), DMD (UBC), MSD, PhD, PPD</td>
<td>Burnaby, BC</td>
</tr>
</tbody>
</table>

^a Advanced Education in General Dentistry
4.2 Development of Questionnaire on Content Validity

Content validation involving SMEs begins with the construction and administration of a questionnaire for content validation (CV) to gather quantitative and qualitative information on the clarity, cultural equivalency and relevance of the scale (Lynn, 1986; Haynes et al., 1995; Grant & Davis, 1997).

I designed a questionnaire with three comprehensive subscales for SMEs to judge: 1) the content validity of OHIP-14K in terms of the technical quality of the items, instructions, and response formats; 2) the relevance of the content to the theoretical domains of Locker’s model; and 3) the cultural equivalency of the translation to the intent of the original OHIP-14 (Appendix D.2). Figure 4 illustrates the scope of each Content Validity subscale with respect to the main stages of the development of OHIP-14K.

On the recommendation of Lynn (1986), the CV questionnaire acquired a 4-point ordinal Likert scale as a response format without neutral ground.

Figure 4. Stages of OHIP-14K development and proposed subscale indices of the content validity questionnaire
4.2.1 Clarity Index

In this part of the CV questionnaire, the SMEs were asked to rate the clarity of the instructions, items, and response format on the following Likert scale: 0 = not at all clear, 1 = somewhat clear, 2 = mostly clear, and 3 = very clear (Kalton & Schuman, 1982; Ferketich, 1991). In the textboxes provided they were instructed to identify comprehension problems they experienced due to vague wording, ambiguous language, complex sentences, double-barrelled questions, dental jargons, and cognitively challenging or culturally inappropriate questions; and to add a comment to explain their response or suggest changes if necessary.

4.2.2 Cultural Equivalence Index

Cultural equivalence implies that the measurement of OHQoL across the two cultures is unbiased. Because all versions of the OHIP, including OHIP-K, are derived from the original in English (Slade & Spencer, 1997), the validity of comparisons between any two cultures primarily depends on the faithfulness and appropriateness of the translation from the source English version (Gagnon & Tuck, 2004). In other words, the original English version is the basis for evaluating the cross-cultural equivalency of the OHIP-14K.

In my study, the original OHIP-14 and the OHIP-14K were collated using expert judgments to examine the cross-cultural equivalence of the two versions. The SMEs were asked to evaluate the semantic, colloquial, experiential and conceptual equivalence, rather than linguistic or literal equivalence of the translation using the scale: 0 = not at all equivalent, 1 = somewhat equivalent, 2 = mostly equivalent, and 3 = equivalent. They were encouraged also to comment generally on the translation with suggestions for deletions, additions and modifications (Appendix D.2).
4.2.3 Relevance Index

Relevance was determined using expert ratings on the relevance index, which gauged the degree to which the OHIP-14K items appropriately sample the theoretical domains of OHIP. As suggested by John and colleagues (2005), the bilingual SMEs were asked to identify for each item the most appropriate theoretical domain of Locker’s model: functional limitation, physical pain, psychological discomfort, physical disability, psychological disability, social disability or handicap. The chosen response method on the relevance index was a multiple-choice question derived from the Q-Sort and inter-rater agreement methods. In Q-sort selection method, experts are asked to place index cards containing items into the most appropriate categories (Block, 1961; Waltz & Bausell, 1991). This method has been widely used to refine items and verify that developed items would correctly load onto expected domains at a pretesting stage (Vaughn & Waters, 1990; Moore & Benbasat, 1991; Bhattacherjee, 2002; Nahm et al., 2002; van Ijzendoorn et al., 2004), which in turn forms the basis of construct validity and improves the reliability of the constructs (Rajesh et al., 2010). However, the Q-sort method I used was modified because the original method would have required the SMEs’ presence during the course of the validation procedure (Tojib & Sugianto, 2006). To overcome this limitation, I combined the Q-Sort with inter-observer agreement to create a multiple-choice response format for measuring the proportion of SMEs who correctly classified an item to its expected domain (Thorn & Deitz, 1989). If a SME felt that an item did not describe any of the seven domains provided, they were asked to suggest a more appropriate domain in the provided textbox. Finally, the SMEs were asked to write their comments and overall opinions of OHIP-14K in open-ended textboxes.
4.3 Administration of Content Validation Questionnaire and Data Collection

Prior to its administration, the CV questionnaire was pilot tested for clarity by one of the 10 SMEs who recommended that the WHO’s definitions of impairment, disability, and handicap be appended to the questionnaire as background information for the other participants. All information in the cover letters, including the various types of cultural equivalence (e.g., semantic equivalence), were verbally conveyed to the SMEs. After this introductory briefing, the SMEs signed the informed consent form (Appendix D.1) and received a CV questionnaire with an instruction manual and the OHIP-14 in both Korean and English (Appendix D.2). They answered the CVI questionnaires individually, independently, and at their own convenience. I returned within 30 days to collect the questionnaire. During this 30-day period, I telephoned each SME every two weeks to prompt their response, and when I finally collected the completed questionnaire, I had a brief discussion with each of them about their opinions on the process and offered them an honorarium of $50 for their participation.

4.4 Data Analysis

The Likert responses were transferred to and analyzed using Microsoft Office Excel®. Each scale was inspected for any missing responses, and the mean score for a missing response was imputed for the mean score for that scale element if the missing data did not exceed 60% and items with missing data did not exceed 15% of the total number of items (Fox-Wasylyshyn & El-Masri, 2005; Meulen et al., 2008; Sanders et al., 2009).

The clarity of the OHIP-14K was rated for relevance and equivalence both globally for the entire scale (S-CVI) and individually for the instructions, response format, event frequency, and items (I-CVI).
4.4.1 Calculating the Item-level Content Validation Index (I-CVI)

The I-CVIs were calculated as the proportion of SMEs who endorsed the validity of each scale (i.e., gave ratings of 3 or 4 on the 4-point Likert scale) (Beck & Gable, 2001).

4.4.1.1 Calculating the Average Deviation Mean Index

$AD_M$ is an index of disagreement with the ability to measure multirater disagreement in the scale’s units (Burke & Dunlap, 2002) and to uncover hidden disagreement in dichotomous data (i.e., content valid vs. content invalid). $AD_M$ for the clarity and cultural equivalence indices was calculated as the sum of the differences between individual ratings and the mean in absolute values divided by the total number of ratings. A lower $AD_M$ value indicated stronger agreement between SME-ratings because the $AD_M$ is dispersed around the mean. The detailed statistical formulas for calculating $AD_M$ for an item and a scale are provided in Appendices C.1 and C.2.

Two cut-off values of $AD_M$ were used to interpret agreement and disagreement between SMEs. The minimum acceptable level of disagreement ($AD_M$) was calculated as $c/6$, where $c$ is the number of categories (i.e., 4), which was 0.67 for the 4-point Likert scales (Lynn, 1986). An $AD_M$ value of 0.67 or higher indicated a *significant level of disagreement* among SMEs (Figure 5). Any scale element with an $AD_M$ greater than this cut-off value required further inquiry into the opinions and explanations of the SMEs to identify alternative questions or format. On the other hand, a critical $AD_M$ value of 0.56 or lower was required for the 5% level of *statistically significant agreement* (achieving the 5% level of statistical significance) for a study involving 10 experts and 4 response categories (Burket & Dunlap, 2002). $AD_M$ values
below this cut-off value indicate that agreement was unlikely to have been achieved by chance, and it was necessary to accept the SMEs’ feedbacks and suggestions for making revisions. ADM values between 0.56 and 0.67 indicated that neither significant disagreement nor agreement had been reached, and that no further action was required.

Figure 5. Possible outcomes of content validation study of OHIP-14K
4.4.1.2 Multirater Kappa

The analysis of nominal or categorical data generated by the relevance index in which the most representative theoretical domain was selected for each item requires non-parametric Kappa statistics (Slocumb & Cole, 1991). The minimally acceptable Kappa cut-off value was 0.4 (Okochi et al., 2005; Gorelick et al., 2008). Kappa values below this cut-off were considered to be poor agreement (Fleiss, 1971; Cicchetti, 1984) and prompted further examination of the SMEs’ comments. Kappa values at and above 0.40 indicated moderate agreement. Detailed calculations and interpretation of free-margin $K_{free}$ are provided in Appendices C.2 and C.3.

4.4.2 Scale-Content Validity Index

The Scale-level CVI, expressed as the percentage of items whose I-CVI values were equal to or greater than the minimally acceptable CVI of 0.78, provided information on the proportion of elements requiring revisions until the S-CVIs was equal to or greater than 0.80 to confirm the validity of the scale (Grant & Davis, 1997).
Chapter 5: Results

5.1 Clarity of OHIP-14K Elements

With the exception of the response format, all 16 elements (including items and instructions) had I-CVI values greater than the critical value of 0.80, indicating adequate levels of clarity (Table 5). The deviation from the mean index \((\text{AD}_M)\) across all elements was below the critical value of 0.56, suggesting that homogeneity in SME ratings was unlikely to have been achieved by chance.

For the most part, the OHIP-14K was unequivocally judged to be clear \((\text{S-CVI} = 0.93, \text{AD}_M \leq 0.56)\). The only scale element whose CVI value fell below the acceptable level was the response format \((\text{I-CVI} = 0.7)\). The comments from three SMEs suggested that the response format was vague. They recommended changing the Korean words “maewoo” (very) to “maewoo jaju” (very often), and “guhee” (hardly) to “guhee junhyu” (hardly ever). In addition, three other SMEs commented that the Korean translation of the OHIP-14 asks about symptoms that overlap significantly with systemic illnesses such as depression, so it could potentially have diverse interpretations. They recommended that the specific oral health context be expressed in every question by including the phrase “because of problems of your mouth, teeth or dentures.”
Table 5. Item-level content Validity Index (I-CVI) and Average Deviation from the Mean Index (ADM) of OHIP-14K instruction, response format, event frequency, and items in the Clarity and Equivalence Indices

<table>
<thead>
<tr>
<th>Original Scale Element (back-translation of OHIP14-K)</th>
<th>Clarity Index</th>
<th>Equivalence Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I-CVI</td>
<td>ADM</td>
</tr>
<tr>
<td>Instruction</td>
<td>0.9</td>
<td>0.38</td>
</tr>
<tr>
<td>Response Format</td>
<td>0.7</td>
<td>0.34</td>
</tr>
<tr>
<td>Event Frequency&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.0</td>
<td>0.40</td>
</tr>
<tr>
<td>1. Trouble pronouncing any words</td>
<td>1.0</td>
<td>0.40</td>
</tr>
<tr>
<td>(Discomfort from not being able to pronounce well)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Sense of taste has worsened</td>
<td>1.0</td>
<td>0.38</td>
</tr>
<tr>
<td>(Sense of taste was worse than before)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Painful aching in your mouth</td>
<td>1.0</td>
<td>0.40</td>
</tr>
<tr>
<td>(Pain in the tongue, sublingual, cheeks, palate, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Uncomfortable to eat any foods</td>
<td>1.0</td>
<td>0.39</td>
</tr>
<tr>
<td>(Uncomfortable to have meal due to painful or uneasy problems of the mouth)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Self-conscious</td>
<td>0.9</td>
<td>0.38</td>
</tr>
<tr>
<td>(Reluctant to meet others because of shame)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Felt tense</td>
<td>0.9</td>
<td>0.38</td>
</tr>
<tr>
<td>(Paid attention to)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Unsatisfactory diet</td>
<td>1.0</td>
<td>0.40</td>
</tr>
<tr>
<td>(Dissatisfied meals)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Meals interrupted</td>
<td>1.0</td>
<td>0.40</td>
</tr>
<tr>
<td>(Interrupted during meals)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Difficult to relax</td>
<td>1.0</td>
<td>0.39</td>
</tr>
<tr>
<td>(Difficulties resting comfortably)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. A bit embarrassed</td>
<td>0.9</td>
<td>0.39</td>
</tr>
<tr>
<td>(Embarrassed or perplexed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. A bit irritable with other people</td>
<td>0.9</td>
<td>0.37</td>
</tr>
<tr>
<td>(Get angry easily at others)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Difficulty doing your usual jobs</td>
<td>0.9</td>
<td>0.37</td>
</tr>
<tr>
<td>(Difficult to do normal jobs)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Two missing values from two questionnaires (SME 7 and 8) were imputed with the item mean value for that scale element.
5.2 Cultural Equivalence between OHIP-14 and OHIP-14K

The SMEs were instructed to evaluate the cross-cultural equivalence between the OHIP-14E and OHIP-14K. Seven elements – which included the response format as well as items 7, 8, 9, 10, 12, and 13 – were deemed content valid (CVI > .80), all with acceptable agreement (AD_M ≤ 0.56). On the contrary, the instructions, the event frequency, and items 1, 2, 3, 4, 5, 6, 11, and 14 fell below the minimally acceptable CVI value. Of these scale elements, the instructions, event frequencies, and items 5 and 6 showed statistically significant agreement (I-CVI < 0.80, AD_M ≤ 0.56).

While neither significant agreement nor disagreement was observed for items 2 and 4, a high level of disagreement was noted for items 1, 3, 11, and 14 with regard to cultural equivalency.

- The SMEs were divided over the cultural equivalency of the Korean translation of item 1 trouble pronouncing words (AD_M = 0.96). Four SMEs suggested that having trouble pronouncing words has a different meaning than the Korean phrase discomfort from not being able to pronounce well. The suggested revisions for item 1 included “baleumeul mothaeshinjuk” (could not pronounce [any words]) and “baleumeul mothaesuh himdeushinjuk” (having difficulties to pronounce [any words]).
- In addition, 5 SMEs pointed out that the expression “venting anger at others” (item 11) in Korean is not equivalent to the English “a bit of irritability with other people.” Two SMEs offered an alternative expression, “yakganeu jjajeungeul jal naenjeok,” meaning “showing a little bit of irritability.”

- Lastly, the experts disagreed over the equivalency of item 14 (ADM = 0.72), “totally unable to function,” which was translated into “jungshinjeok, shinchejeok, sahoejeokeuro junhyuh jemokeul halsu upsutdun jeok” (totally unable to do one’s share psychologically, physically, and socially).

5.3 Relevance of OHIP-14K Domains

Table 6 shows the distribution of responses to the question: “Which one of the domains best represents each item?” and a table of descriptive statistics including each item’s CVI value, $K_{free}$, and levels of agreement.
Table 6. The SMEs’ classification of items into the seven OHIP domains, Item-level Content Validity Index (I-CVI), Free-Marginal Kappa ($K_{free}$), and levels of agreement on the Relevance Index

<table>
<thead>
<tr>
<th>Item</th>
<th>Functional Limitation</th>
<th>Physical Pain</th>
<th>Psychological Discomfort</th>
<th>Physical Disability</th>
<th>Psychological Disability</th>
<th>Social Disability</th>
<th>Handicap</th>
<th>I-CVI</th>
<th>$K_{free}$</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>8</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td></td>
<td>0.80</td>
<td>0.56</td>
<td>Moderate</td>
</tr>
<tr>
<td>Q2</td>
<td>9</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.90</td>
<td>0.77</td>
<td>Substantial</td>
</tr>
<tr>
<td>Q3</td>
<td>-</td>
<td>10</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>Perfect</td>
</tr>
<tr>
<td>Q4</td>
<td>6</td>
<td>-</td>
<td>3</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.00</td>
<td>0.30</td>
<td>Fair</td>
</tr>
<tr>
<td>Q5</td>
<td>-</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td></td>
<td>0.50</td>
<td>0.35</td>
<td>Fair</td>
</tr>
<tr>
<td>Q6</td>
<td>-</td>
<td>-</td>
<td>10</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>Perfect</td>
</tr>
<tr>
<td>Q7</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>-</td>
<td></td>
<td>0.40</td>
<td>0.17</td>
<td>Poor</td>
</tr>
<tr>
<td>Q8</td>
<td>2</td>
<td>-</td>
<td>2</td>
<td>6</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.60</td>
<td>0.27</td>
<td>Fair</td>
</tr>
<tr>
<td>Q9</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>8</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.00</td>
<td>0.59</td>
<td>Moderate</td>
</tr>
<tr>
<td>Q10</td>
<td>-</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td>3</td>
<td>2</td>
<td></td>
<td>0.30</td>
<td>0.19</td>
<td>Poor</td>
</tr>
<tr>
<td>Q11</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>8</td>
<td></td>
<td>0.80</td>
<td>0.56</td>
<td>Moderate</td>
</tr>
<tr>
<td>Q12</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>-</td>
<td>2</td>
<td></td>
<td>0.00</td>
<td>0.20</td>
<td>Poor</td>
</tr>
<tr>
<td>Q13</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>-</td>
<td>6</td>
<td>-</td>
<td></td>
<td>0.00</td>
<td>0.38</td>
<td>Fair</td>
</tr>
<tr>
<td>Q14</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>9</td>
<td></td>
<td>0.90</td>
<td>0.77</td>
<td>Substantial</td>
</tr>
</tbody>
</table>

*Note:* Shaded cells indicate the theoretical domains predetermined by Slade and Spencer (1994).

Examination of each item’s relevance to the expected theoretical domain revealed that the endorsement rates for items 1, 2, 3, 6, 11, and 14 were equal to or above the acceptable CVI
value of 0.80 (S-CVI\textsubscript{relevance} = 0.42, $K_{\text{free}} > 0.4$), confirming that the questions correctly corresponded with the domains hypothesized by the Locker’s model. On the other hand, eight out of 14 items (numbers 4, 5, 7, 8, 9, 10, 12, and 13) fell below the acceptable CVI value, with 5 showing poor or fair agreement ($K_{\text{free}} < 0.4$). Closer examination of content-invalid items revealed that item 5, “self-conscious” (psychological discomfort); item 9, “difficult to relax” (psychological disability); and item 13, “life in general was less satisfying” (handicap) were also representing social disability ($K_{\text{free}} = 0.4$), physical disability ($K_{\text{free}} = 0.6$), and psychological disability ($K_{\text{free}} = 0.4$), respectively. Poor agreement was noted for items 7, 10, and 12 ($K_{\text{free}} < 0.2$) while items 4 and 8 showed fair agreement (0.2 < $K_{\text{free}}< 0.4$).

Overall, 7 SMEs indicated that some questions could be interpreted outside of the oral health context. The SMEs recommended specific examples of painful areas (e.g., mouth) and clarifying the oral health context by including a phrase as in the English version: “because of problems with your mouth, teeth or dentures.” For example, item 12, “difficulties in doing one’s usual jobs,” may unintentionally elicit non-dental-related experiences.
5.4  Recommended Revisions or Deletions

A number of potential sources of biases emerged from the content validation results, along with comments and suggested revisions. The potential biases that achieved acceptable agreement are described below within the framework of equivalence theory: methods, items and construct.

5.4.1  Methods

Method biases associated with discrepancies in the instructions and response formats between the OHIP-14E and OHIP-14K were as follows.

<table>
<thead>
<tr>
<th>OHIP-14E Instruction</th>
<th>How often have you had the problem during the last year?</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHIP-14K Instruction</td>
<td>지난 3개월 동안 치야나 잇몸 때문에 아래의 경험을 얼마나 자주 겪으셨습니까?</td>
</tr>
<tr>
<td></td>
<td>(In past 3 months, have you had the experiences below because of teeth or gums?).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OHIP-14E Response Format</th>
<th>Very often, fairly often, occasionally, hardly ever, never.</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHIP-14K Response Format</td>
<td>매우: 1주일에 2–3회 이상, 자주: 1주일에 1회 정도, 가끔: 한달에 2–3회 정도, 거의: 한달에 1회 이하, 전혀: 경험한 적이 없음 (Very: 2–3 times a week; often: once a week; sometimes: 2–3 times a month; almost: once a month; never: never experienced).</td>
</tr>
</tbody>
</table>

According to one SME, the instructions may pose a bias due to the difference in their recall periods (1 year versus 3 months). Another SME noted that a “don’t know” response option is not offered in OHIP-14K.
5.4.2 Items

- Semantic bias – in item 6, the Korean translation of the word *tense* has a different meaning than the English word.

**OHIP-14E Item 6** Have you felt *tense* because of problems with your teeth, mouth or dentures?

**OHIP-14K Item 6** “Shingyungee mani sseuin jeokee itseupnikka?” (Have you paid a lot of attention [to the problem]?).

**Suggested Revision** “Chia, gugangina, teulni taimunae kinjangee doeshinjeokee itseupnikka?” (Have you been tense because of problems with your teeth, mouth or dentures?).

- Functional Bias – although item 9 was perceived to be cross-culturally equivalent (CVI\textsubscript{equivalence} = 0.8, AD\textsubscript{M} = 0.36), 8 of the SMEs recognized *resting comfortably* as a physical disability instead of placing it in the intended psychological disability domain (CVI\textsubscript{relevance} = 0.0, K\textsubscript{free} = 0.59). One SME suggested a revision, replacing the item with “*kinjangeul pulji mothashinjeoki*” to describe the *inability to ease tension*.

**OHIP-14E Item 9** Have you found it *difficult to relax* because of problems with your teeth, mouth or dentures?

**OHIP-14K Item 9** “Pyeonanhage swiji mothashinjeoki itseupnikka?” (Have you had difficulties resting comfortably?).

**Suggested revision** “Chia, gugangina, teulni taimunae kinjangeul pulji mothashinjeoki itseupnikka?” (Have you been unable to ease tension because of your teeth, mouth or dentures?).
• Conceptual bias – OHIP-14K item 5 (I-CVIEquivalence = 0.4), which asks about psychological discomfort, is different than its English counterpart both in meaning as well as the domain covered. This finding is further supported by half of the SMEs classifying item 5 as a social disability (I-CVIRElevance = 0.5) instead of placing it in the intended psychological discomfort domain.

<table>
<thead>
<tr>
<th>OHIP-14E Item 5</th>
<th>OHIP-14K Item 5</th>
<th>Suggested revision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you been self-conscious because of your teeth, mouth or dentures?</td>
<td>“Changpihaeseo dareun sarameul mannagiga kkeoryeojishin jeoki itseupnikka?” (Have you been reluctant to meet others because of embarrassment?).</td>
<td>“Chia, gugangina, teulni taimunae shingyeongee mani sseuishinjeoki itseupnikka?” (Have you paid attention [to the problem] because of your teeth, mouth or dentures?).</td>
</tr>
</tbody>
</table>

5.4.3 Construct

• Construct bias – the seemingly transferrable construct of OHQoL can be understood differently in English and Korean cultures due to differences in priorities, health perceptions and potential impact of a disorder. For example, one SME indicated that OHIP-14K does not adequately capture the aesthetic concerns that patients might have about their mouths. In this regard, item 22 of the long form OHIP-49 (appearance unsatisfied) might be more relevant to Koreans (Bae et al., 2007). In their overall evaluation of OHIP-14K, 6 SMEs recognized the need for better accuracy of the OHIP-14K items to ensure cross-cultural equivalence.
Chapter 6: Discussion

The availability of cross-culturally valid and reliable OHQoL measures is beneficial to Korean populations for needs assessment, oral health care planning treatment and outcome and service evaluation. Despite the widespread use of OHIP, a review of the literature revealed that content validity and equivalency of its translated versions – including the OHIP-14K – are rarely addressed. This was compounded by the fact that current cultural adaptation and validation strategies are not resistant to biases that can have serious ramifications to inferences made on cultural differences in OHQoL. Typical validation efforts for the OHIP use criterion-related approaches that are vulnerable to the cross-cultural biases and misunderstandings. However, they have paid little attention to the content of the scale or theoretical foundations. Consequently, the validity of the OHIP translated to other languages for use in cultural groups other than English-speakers in Western countries is suspect.

This literature finding prompted an investigation of content validity and cultural equivalency of OHIP-14K in order to support the construct validity and equivalence of OHQoL measurements in Korean. I was also interested in applying the CVI techniques for detecting and addressing potential cross-cultural biases. My thesis began with a literature review to provide an overview of the contemporary conceptualization of OHQoL, validity, and cultural equivalence, followed by a literature synthesis of existing approaches to adapting and validating the OHIP. The second part of my study involved a content validation of OHIP-14K with bilingual Korean SMEs.

A number of studies have looked at the content validity of translated versions of psychometric scales. The authors of the Dutch version of the Safety Attitudes Questionnaire used CVIs and Kappa to assess the equivalency with the English version (Devriendt et al., 2012), while Shiu
and colleagues (2003) used a similar method on the Chinese version of the Diabetes Empowerment Scale. Both studies concluded that the translations were equivalent to the original, although they did not use bilingual subjects to help with the assessments despite evidence suggesting that bilingual experts yield more effective translation (Guillemin, 1995; Harkness et al., 2003). In contrast to the traditional committee method of establishing equivalency, my study investigated the theoretical foundations of the OHIP-14K and recommended the structured methods for resolving potential biases. In the following 4 sections, I will discuss the method of content validation used, content validation findings, possible implications of my work and study limitations, while tying each of these to my literature review.
6.1 Content Validation Methods

Content validation studies in nursing, education, and exercise physiology researches used different numbers of SMEs (Chen et al., 2003; Chien & Norman, 2004; Li & Lopez, 2004; Hubley & Palepu, 2007; Devriendt, 2012), but the CVI techniques and indices they used were similar to my method. Chen and colleagues (2003) used a 4-point Likert scales to measure the clarity of a Chinese translation of the Tobacco Acquisition Questionnaire and Decisional Balance Scale with results (S-CVI = 0.96) that were close to my S-CVI of 0.93.

The cultural equivalence index, which I adapted from Lynn’s CVI (1986) for appraising the item-level equivalency of OHIP-14K, already existed as the Translation Validation Index, which was also derived from the CVI (Tang & Dixon, 2002). It has been applied also to four different questionnaires translated to Japanese (Kochinda, 2011) and a Chinese version of the Adolescent Teasing Scale (Liu, 2011). Like my study, they all used the same cut-off value of 0.80 for the cultural equivalence index. Similar to my results (CVI_{equivalence} = 0.50), the Japanese versions of the Emotional Response Questionnaire (TVI = 0.46), the Style of Helping Scale (TVI = 0.73), and the Threat-Support Scale (TVI = 0.73) raised serious questions about the equivalency of the translated scales. On the other hand, the Japanese version of the Intensity of Help Scale (TVI = 0.80) and the Chinese version of the Adolescent Teasing Scale (TVI = 0.88) fared much better.

Domingues and colleagues (2011) also used a 4-point scale to demonstrate the semantic-idiomatic, cultural and metabolic equivalence of the Brazilian translation of the Veterans Specific Activity Questionnaire (VSAQ). Unlike many other self-reported health measures, VSAQ is a clinical tool used to measure exercise levels and is not grounded in a theoretical model (Myers et al., 1994). The OHIP-14K, in contrast, is based on Locker’s model of oral
health to sample seven specific domains. Therefore, the relevance of the OHIP-14K to Locker’s model must be considered in any assessment of equivalency between different translations.

In the literature, relevance indices were used by Chien and Norman (2004), who employed 15 bilingual health professionals to determine the appropriateness of the Chinese version of the Family Burden Interview Schedule in addressing the original dimensions of the English version (S-CVI = 0.96) (2004). In a similar study by Li and Lopez, 10 nursing experts evaluated the relevance of the Chinese version of the State Anxiety Scale for Children and also reported a very high score (S-CVI = 0.96) (2004). Both studies showed a substantially higher CVIs than that obtained in this study (S-CVI_{relevance} = 0.42), suggesting that the low values of relevance for OHIP-14K may not be related to the difficulties of the rating tasks.

By comparing these different CVI methods, it is apparent that the indices of clarity, translation equivalence and relevance used in my study provides a thorough appraisal of the OHIP-14K. However, I employed these indices in a slightly different manner in terms of response format on the relevance index and the use of disagreement indices to control chance agreement, as I explain below.

### 6.1.1 Response Format on the Relevance Subscale

The traditional relevance indices used by Chien and Norman (2004) and Li and Lopez (2004) measured how relevant an item is to its theoretical domain. Such an arrangement, however, can bias the SMEs’ judgments by exposing them to a predetermined item-domain association, and it could still remain unclear which dimensions the questions are addressing. For these reasons, the relevance index I used had a multiple-choice format derived from the Q-sort method and hybridized with inter-observer agreement. Slocumb and Cole (1991) demonstrated
the utility of this categorization method by which SMEs were given definitions of five theoretical domains (Universal, Competence, Availability, Resource, and Embracing) and asked to assign each item to a domain or identify it as independent of all of them. Rubio and colleagues (2003) referred to this method as factorial validity index (FVI) and used it to determine the extent to which experts correctly identify items with particular domain. The authors recommended a minimally acceptable FVI value of 0.8, which was consistent with my study design. However, the FVI is plagued also by chance agreement.

6.1.2 Disagreement Indices

The traditional content validation studies were often criticized for running the risk of possible chance agreement (Cohen, 1960; Polit & Beck, 2006; Tojib & Sugianto, 2006). Apparently, a proportion of agreement indices can both over- and under-estimate agreement without a correction for coincidental agreement and disagreement (Burke et al., 1999). Lynn (1986) developed a minimally acceptable range of agreement by calculating a standard error of proportions with a 95% confidence interval around each I-CVI value. Consequently, I selected a minimum of 8 out of 10 SMEs required to establish content validity beyond the 5% significance level (I-CVI = 0.78).

Apart from the risks of chance agreement, the CVI alone cannot account for the loss of data as a result of collapsing 4-point Likert responses into a bivariate content-valid or -invalid choice. To overcome this limitation, I supplemented the CVI assessment with an additional analysis of ADM and multirater Kfree for better accuracy of and confidence in CVI measurement. ADM, a measure of the divergence of responses, has been used to supplement the CVIs of the Injection Drug User Quality of Life measure with a cut-off of 0.69 for acceptable disagreement (Hubley & Palepu 2007). The lower cut-off value (ADM = 0.56) used in my study meant that a lower
A degree of disagreement in expert ratings was required to establish consensus on the CVI results.

On the other hand, I applied free-margin multirater $K_{\text{free}}$ instead of AD$_M$ to the data obtained from the relevance index and interpreted the $K_{\text{free}}$ values with parameters provided by Landis and Koch (1977). In contrast to the CVIs, Kappa is compared to the maximum possible value without chance agreement (Musch et al., 1984). Stemming from the classical test theory, the Kappa statistic was originally developed by Cohen (1960) to correct for chance agreement and disagreement between two judges. Fleiss (1971) extended the Kappa to allow for classification of the non-numerical variables by more than two judges. However, Fleiss’s Kappa was inappropriate for my study because it has fixed-marginal ratings with a predetermined number of items belonging to each domain. Instead, I used multirater $K_{\text{free}}$ to analyze data gathered from SMEs who were not restricted to assigning a certain number of items to each domain (Randolph, 2005). Wynd and colleagues (2003) used this combination of CVI and multirater $K_{\text{free}}$ to assess experts’ agreement on the content validity of the Atkins Osteoporosis Risk Assessment Tool (ORAT), which had an overall CVI of 0.65 and poor agreement (0.29 ≤ $K_{\text{free}}$ ≤ 0.48). Chung and colleagues (2007) also used the same method with similar cut-off values (I-CVI > 0.7, $K_{\text{free}}$ > 0.4) to examine the content validity of the Chinese version of the Integrative Medicine Attitude Questionnaire and found limited evidence of content validity (S-CVI = 0.71) and expert agreement ($K_{\text{free}}$ = 0.09). Similarly, I found a low level of relevance for the OHIP-14K (S-CVI$_{\text{relevance}}$ = 0.42) with varying degrees of agreement between SMEs (0.17 ≤ $K_{\text{free}}$ ≤ 1.00).

Another notable difference between my study and other content validation studies is that the cultural adaptation of OHIP-14K was not guided by a translation theory or model. For example,
many cross-cultural studies in nursing adhere to a standardized set of procedures called the *state-of-art method of translation and validation* (Brack & Barona, 1991; Chien & Norman, 2004; Li & Lopez, 2004). There are also other translation models (Guillemin, 1995; Beaton et al., 2002) characterized by a synthesis of multiple independent translations so as to ensure the quality, faithfulness, and cultural appropriateness of an adaptation. However, my systematic synthesis provided insights into its guiding principles and allowed also for an appraisal of the methods. The results of my synthesis showed that cultural adaptation of OHIP is in general guided by a methodological norm in an ad-hoc fashion rather than by a specific model.

Additionally, I found three different S-CVI calculation methods in the literature: universal agreement (S-CVI/UA) (Rubio et al., 2003); averaging method (S-CVI/Avg) (Polit & Beck, 2006); and proportional S-CVI (Lynn, 1986). S-CVI/UA represents the proportion of elements that achieved ratings of 3 or 4 by all the SMEs. Because unanimous agreement by multiple experts is difficult to achieve, Rubio and colleagues (2003) proposed averaging all I-CVIs on the index in order to calculate S-CVI/Avg. However, averaged I-CVI values do not give us any information about what percentage of the items are content valid. Therefore, I used the proportional S-CVI method put forward by Lynn (1986).

In summary, my content validation methods were compared and contrasted with those found in the literature. Now I will turn to the discussion of my findings and their implications for cross-cultural measurement and comparison of OHQoL.

### 6.2 Discussion of Content Validation Findings

Within CTT, the obtained CVIs can be interpreted as a combination of true scores and random errors in the experts’ ratings. Measurement errors in content validity are thought to have a mean of zero and do not correlated with either the true scores or with each other (Lord &
Novick, 1968). The true scores in the CVIs could be due to the following reasons (Murphy & De Shon, 2000):

1. Shared perceptions of content validity and cultural equivalency.
2. Shared goals and biases (i.e., rater influenced in the direction of leniency).
3. Shared frames of reference (similar standards or expectations).
4. Shared relationships (acculturation of bilingual SMEs).

The scale-level CVIs obtained in my study indicated that wording of the OHIP-14K is clear (S-CVI\textsubscript{clarity} = 0.93), but it is not cross-culturally equivalent to its English counterpart (S-CVI\textsubscript{equivalence} = 0.50). The results on the clarity index are expected, considering that OHIP-14K has already undergone rigorous pilot testing with monolingual Korean adults and five Korean dentists (Bae et al., 2007) despite its cultural equivalency never having been fully established.

Moreover, my results seemed to indicate a limited degree of relevance for the entire scale (S-CVI\textsubscript{relevance} = 0.42). Cross-culturally valid scales must demonstrate evidence of item relevance and mutual exclusiveness of their theoretical dimensions to achieve proper representation of the construct (Smit et al., 2006). For this reason, many researchers recommend that each domain be represented by at least two content-valid items as a safety net to avoid taking chances with translation quality (Bice & Kalimo, 1971; Hinkin et al., 1992). In my study, only 7 out of 14 OHIP-14K items accurately loaded onto the expected dimensions. The low level of item relevance can imply a departure from the original theoretical model and an inaccurate representation of the construct since the subscales are not measuring what they are supposed to measure. This finding also raises questions about the use of subscale scores as a valid and reliable indicator of OHQoL domains.
Contrary to my expectations, a significant disagreement between SMEs was noted on the relevance and cultural equivalence indices. According to CTT, highly correlated ratings reflect the true score while observed disagreement is attributed to random measurement errors (Lord & Novick, 1968). In my study, however, the high levels of disagreement could also indicate a failure to achieve a shared understanding of the theoretical construct or dimensions as well as individual differences in interpretation of oral problems and oral health behaviours. In fact, Murphy and De Shon (2000) offered additional explanations for observed disagreement among SMEs, which might also have been the case in my study.

Potential causes of multirater disagreement
1. Systematic differences in their ratings of CVI and cultural equivalency.
2. Systematic differences in areas or levels of expertise or linguistic proficiency.
3. Systematic differences in access to information other than the scale provided (e.g., previous exposure to OHIP).
4. Individual differences in the interpretation of words or expressions with multiple meanings.
5. Individual differences in background, rating proficiency, clinical knowledge and experience, levels of expertise and education.

Scale elements with high disagreement required further inquiry into the SMEs’ comments for possible explanations. Following Winderfelt’s suggestions (2005), the qualitative inputs from the SMEs for the content-invalid items with AD_M ratings of 0.67 or higher were recorded to make alternative terms available for further validation studies.

On the other hand, the scale elements meeting the minimally acceptable disagreement level (AD_M \leq 0.56 or K_{free} > 0.40) but falling below the acceptable value of CVI (0.8) needed to be revised or eliminated according to the experts’ suggestions. Instead of eliminating items from
a scale like OHIP-14, which is already short, my study findings suggest revisions of the OHIP-14K instructions, response format and frequency as well as items 5, 6, and 9. Suggested modifications of the OHIP-14K are presented in Table 7.

Table 7. Suggested revisions of OHIP-14K

<table>
<thead>
<tr>
<th>Scale Element</th>
<th>OHIP-14K</th>
<th>Suggested Revisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction</td>
<td>3 months</td>
<td>1-year recall period</td>
</tr>
<tr>
<td>Response Format</td>
<td>“maewoo” (very) and “guhee” (hardly)</td>
<td>“maewoo jaju” (very often) and “guhee junhya” (hardly ever)</td>
</tr>
<tr>
<td>Q. 5</td>
<td>“Changpihaeseo dareun sarameul mannagiga kkeoryeojishin jeoki itseupnikka?” (Have you been reluctant to meet others because of embarrassment?)</td>
<td>“Chia, gugangina, teulni taimunae shingyeonge mani sseuishinjeoki itseupnikka?” (Have you been self-conscious because of problems with your teeth, mouth or dentures?)</td>
</tr>
<tr>
<td>Q. 6</td>
<td>“Shingyungee mani sseuin jeoki itseupnikka?” (Have you paid a lot of attention [to the problem]?)</td>
<td>“Chia, gugangina, teulni taimunae kinjangee doeshinjeok itseupnikka?” (Have you been tense because of problems with your teeth, mouth or dentures?)</td>
</tr>
<tr>
<td>Q. 9</td>
<td>“Pyeonanhage swiji mothashinjeoki itseupnikka?” (Have you had difficulties resting comfortably?)</td>
<td>“Chia, gugangina, teulni taimunae kinjangeul pulji mothashinjeoki itseupnikka?” (Have you been unable to ease tension because of problems with your teeth, mouth or dentures?)</td>
</tr>
</tbody>
</table>

As previously discussed, standardizing uniform scale features, such as recall periods, is necessary to minimize construct-irrelevant variance for making valid cross-cultural comparisons. In addition, the inclusion of the phrase “because of problems with your teeth, mouth or dentures” for all items would specify the affected areas and help readers to focus on oral health-related events when answering the questions. Finally, the content validity findings were presented in English for non-Korean readers of my thesis using online translator tools for
informational purposes only (Brondani & He, 2012). Back-translations of translated items and SME comments are commonly used for presenting content validation study results, as was done by Noh and colleagues when presenting their content validity evidence for the Korean version of the Center for Epidemiological Studies Depression Scale (1992).

In the literature, various types of biases have been reported in other language versions of the OHIP. When item 3, “have you had any painful spots or areas in your mouth?” was translated into Brazilian Portuguese, it used the Portuguese word “pontos” for spot. However, the word also has a second meaning – suture stitches – which caused confusion. In Chinese, question 4 about “meal interruption” was translated as “stop eating during meals” (Wong et al., 2002). Although these findings were not replicated in my study, a similar translation problem to the one reported by Kenig and Nikolovska (2012) in the Macedonian version of the OHIP was detected in item 5, “self-consciousness,” of OHIP-14K. In Macedonian, the literal translation of the term had a different meaning than intended whereas in the Korean version, the low degrees of translation equivalence (I-CVI = 0.4) and relevance (I-CVI = 0.5) suggest that the Korean translation may have been too liberal. A similar semantic problem was reported with item 5 in the Macedonian version. In this case, the authors decided to eliminate the item because most respondents did not understand its meaning (Segu et al., 2005).

There are a number of possible explanations for the reported translation problems. As previously discussed, one possible cause of the observed asymmetry between source and target texts could be the disrupted balance between cultural appropriateness vs. faithfulness in the cultural adaptation. However, striking a balance between the two is a difficult process because a translator’s interpretation of the original concept will not be bias-free, especially without using the original theoretical model as a guiding principle in the cultural adaptation process. In a survey of translators, Harkness and colleagues (2003) found that in absence of adequate task
specifications, translators are forced to provide their own interpretations, which can bias the translation outcomes (1996). This subjective nature of cultural adaptation was also noted by van Ommeren and colleagues (1999), who found the translator’s judgment to be the determining factor in the success of cultural adaptations. Thus, the aim of the cultural adaptation must be clearly communicated to translators in order to carefully define the confines and fluidity of meaning as well as the range of interpretation to be considered (Wilss, 1996). To address issues with the personal biases of translators, Kleiner and colleagues (2009) provided them with instructional guidelines that would help them to produce more faithful and culturally appropriate translations of the National Survey of Children with Special Health Care Needs in Chinese, French and Spanish, and they found a resultant improvement in the quality of translation for some of these languages (2009). In a similar approach, I provided SMEs with the theoretical blueprint of OHIP (i.e., Locker’s model and domain definitions) to help them evaluate the OHIP-14K content against the theoretical model.

Another source of asymmetry could have been the ambiguities in the source English version itself. As discussed previously, the scale’s “handicap” domain included questions that were not developed from the interviews with the lay Australian respondents (Slade & Spencer, 1997). In the case of my study, none of the SMEs judged item 13, “life less satisfying,” to represent the handicap domain. Moreover, high disagreement was noted for item 14, “totally unable to function,” which was translated into “jungshinjeok, shinchejeok, sahoejeokeuro junhyuh jemokeul halsu upsutdun jeok.” One SME pointed out that the Korean translation means a state of being incapacitated psychologically, physically and socially. Not only can such a triple-barrelled question be confusing for respondents, but it also gives a very vague impression of “handicap,” for which no equivalent word exists in Korean. Another SME suggested replacing the wording with “jaeneunggreokeul mothan jeok” (unable to use one’s abilities).
Finally, there could have been conceptual differences across cultures in interpreting the semantically equivalent texts. My study found that seemingly equivalent items did not always guarantee their relevance to the expected theoretical domains. For example, although item 9, “have you had difficulties relaxing?” was judged to be equivalent to its English counterpart (I-CVI\textsubscript{equivalence} = 0.8), it did not load onto the expected psychological disability dimension (I-CVI\textsubscript{relevance} = 0.0). A similar translation issue was reported by Room and colleagues (1996), who noted that Korean translations of psychological affective states were easily mistaken for physical states because Korean words regarding feelings do not effectively differentiate between physical sensations and emotions. Other researchers also noted that semantically equivalent items are not always conceptually equivalent (Hunt, 1986; Guillemin et al., 1993). For instance, semantically equivalent translations of “I have pain in my head” might vary in their significance and severity depending on the culture (Hunt, 1996). In relation to such conceptual and semantic discrepancies, Bice and Kalimo (1971) summarized 4 possible translation outcomes:

1) Identical items are semantically and conceptually equivalent across cultures.

2) Items that are conceptually but not semantically equivalent may still be used in comparative studies without changes (Bice & Kalimo, 1971) unless an alternative mode of expression is available (Hunt, 1986).

3) For culture-linked items that are semantically equivalent but measure different constructs in different settings, reconsideration of the meaningfulness of the responses or revisions may be necessary to establish construct equivalence (Bice & Kalimo, 1971; Hunt, 1986).

4) Non-related items that are neither conceptually nor semantically equivalent must be either re-translated or removed.
Using this typology, the functional (item 9) and construct (item 5) biases that I presented earlier can be classified as culture-linked and non-related items, respectively. Hence, revisions of these items will be critical to the validity of cross-cultural comparisons of OHQoL.
6.3 Implications of the Content Validity Findings

My study found limited evidence of content validity for OHIP-14K in terms of item relevance and cultural equivalency with the source English version. This has further implications for the construct validity of OHQoL measurements in Korean populations using OHIP-14K because in the Unitarian framework, content validity of the scale is considered as the prerequisite for establishing the construct validity of the measurement. A scale with poor content validity can bias the empirical results and can even prompt the acceptance or rejection of a hypothesis or a theory (Rossiter, 2008).

In line with previous research on the integrative framework of emic and etic perspectives (Tripp-Reimer, 1984; Phillips & Luna, 1996; Morris et al., 1999; Alegria et al., 2004; Cenoz & Todeva, 2009), the SMEs’ suggestions for improving the scale’s content validity underscored the importance of its cultural appropriateness and faithfulness to the source version and theoretical mode. This study finding suggests that OHIP-14K should have the same relationship with the construct of interest both within and across cultural groups. Therefore, a theoretical blueprint for the original scale might be useful as a practical baseline for ensuring semantic and conceptual equivalence in cultural adaptation and validation processes. In addition, the further provision of explanations of the scale’s concepts for translators or researchers might be necessary to avoid the conceptual bias caused by misinterpretation of items (Wild et al., 2005).

Another potential implication of my study includes the impact of the detected biases on cross-cultural comparisons of OHQoL. The same degree of construct might elicit different responses on a Likert scale and consequently bias interpretation of domain and total scores (Anderson et al., 1996). This form of bias is called a scalar bias. For example, a physical disability subscale
could be measuring social disability domain instead. The fragile relationship between items and their theoretical domains carries significant implications as OHIP scores are calculated for each domain as well as for the entire scale (Locker et al., 2005; Brondani & MacEntee, 2007).

However, Brondani and MacEntee (2007) raised a more fundamental problem with the use of a summative score, due to the questionable discreteness and stability of the theoretical domains of Locker’s model. Bakers (2007) also questioned whether or not the OHIP domains were distinguishable and if so, how they would relate to one another as per Locker’s conceptual framework. Brennan and Spencer (2004) provided further empirical evidence in support of the existence of conceptual domains using factor analysis, and the 7 domains of Locker’s model were confirmed by John using expert judgment (2007). Therefore, in regard to the aim of the content validity study, the 7 theoretical domains are integral components of Locker’s model that provide a foundation for specifying and operationalizing the construct of OHQoL.

Based on the feedback from the 10 SMEs, a refined version of OHIP-14K was yielded to enhance both semantic and conceptual equivalence (Table 8). However, more work is needed to test the new measure and evaluate its psychometric properties in the target Korean populations (see 7.2. Future Directions).
Table 8. Suggested revised version of OHIP-14K

<table>
<thead>
<tr>
<th>지난 1 년동안 아래의 경험을 얼마나 자주 겪으셨습니까?</th>
<th>매우 자주</th>
<th>자주</th>
<th>가끔</th>
<th>거의 전혀</th>
<th>전혀</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 치아, 구강이나 틀니 때문에 발음이 잘 안되어 불편했던 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. 치아, 구강이나 틀니 때문에 맛을 느끼는 감각이 예전보다 나빠졌다고 느끼신 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. 혈밀, 빨 입천정 등이 아픈 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. 치아, 구강이나 틀니 때문에 아프거나 거북스러운 입안의 문제 때문에 음식 먹기가 불편한 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. 치아, 구강이나 틀니 때문에 신경이 많이 쓰인적가 있습니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. 치아, 구강이나 틀니 때문에긴장이 되신적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. 치아, 구강이나 틀니 때문에식생활이 불만스러운 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. 치아, 구강이나 틀니 때문에 식사를 도중에 중단하신 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. 치아, 구강이나 틀니 때문에 긴장을 풀지 못하신 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. 치아, 구강이나 틀니 때문에 난처하거나 당황스러웠던 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. 치아, 구강이나 틀니 때문에 다른 사람들에게 화를 잘 내게 되신 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. 치아, 구강이나 틀니 때문에 평소 하시던 일을 하기가 어려웠던 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. 치아, 구강이나 틀니 때문에 살아가는 것이 예전에 비해서 덜 만족스럽다고 느끼신 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. 치아, 구강이나 틀니 때문에 정신적 신체적 사회적으로 전혀 제 품을 할 수 없었던 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6.4 Study Limitations

Despite content validity and cultural equivalency being crucial evidence of construct validity for the OHIP-14K, it is important to keep in mind that the findings of my study are based on expert judgment rather than empirical evidence obtained from administering the scale to subject pools. My study was meant to be an adjunct to the mounting evidence gained in the investigation of scale validity from various sources, including the empirical validation of OHIP-14K conducted by Bae and Colleagues (2007). Nonetheless, the findings of the study have three main caveats. The first caveat is related to the literature review and study design. As there was limited literature on content validity assessment for scales measuring OHQoL, the scope of the literature search was expanded to include literature in multiple fields including education, social sciences, business and medicine. Given the breadth of literature and variety of topics covered, the manual literature search might have not been efficient in identifying previous studies that had adopted similar designs under different names while I was formulating my research questions and developing a content validation method. In retrospect, the duplication of my efforts could have been avoided if I had taken a more systematic approach to the literature review.

Furthermore, as with any other study design that uses expert judgment, my content validity study was subjective in nature. Although an individual SME’s error or bias could be addressed by group decisions, collective judgment is not infallible either and may mask disagreement between individual experts. Also, the convenience sample used in my study consisted only of dentists. Multi-disciplinary SMEs could have provided a more diverse range of knowledge and experience (Domingues, 2011; King et al., 2011). Moreover, expert judgment did not allow for further examination of the latent relationships in multiple variables (John, 2007), such as theoretical domains and how they relate to each other.
Another issue pertinent to relying on the expert judgement of bilingual dentists is their bilingual proficiency and the effects of acculturation on the generalizability of their findings. Despite the stringent expert selection criteria, it could not be confirmed whether every SME met the levels of proficiency in both languages and familiarity with both cultures required for the content validation tasks. An individual expert’s linguistic competence and cultural knowledge could potentially have a huge impact on the study results concerning the equivalence and relevance of OHIP-14K. In anticipation of similar problems, I attempted to attenuate the impact of individual biases on the study results by employing percent agreement and disagreement indices.

Apart from the individual biases, the bilingual experts may have had differences in their backgrounds and linguistic processing from the average monolingual Korean layperson. Hence, there is a need to examine how target respondents interpret the scale content since poor understanding of the instructions and questions can affect the accuracy of recall and self-reporting of past experiences (Linn et al., 1991).

The second caveat concerns the particular method of content validation used in my study. While privacy and anonymity in data are the strengths of CVIs for encouraging honest feedback and minimizing the influence of bias and peer pressure from other experts, they are also considered weaknesses in that they limit cooperation between SMEs and minimize accountability for their individual judgments (Goodman, 1987). Researchers are divided over whether anonymous or non-anonymous conditions would provide more accurate ratings. Leek et al. (2011) found that collaboration between referees produced a more accurate peer-review process, but there are also studies that support anonymous evaluation because accountable ratings may suffer a significant level of inflation (Afonso et al., 2005; Daberkow et al., 2005).
More empirical evidence is needed to determine which settings are more appropriate for a content validity assessment.

The other potential drawback of the chosen CVI analysis is related to the use of Kappa statistics for analyzing the data collected on the relevance index. Kappa statistics did not permit examination of the interrelationship between conceptual dimensions; it is designed to analyze nominal data of no natural ordering even though in this case the dimensions were sequentially related, as hypothesized by Locker’s model and ICIDH. For example, one can infer from Locker’s model (Figure 2) that impairment would be more closely related to functional limitation than to physical disability.

The third caveat of my study design pertains to the limitations inherent in the original OHIP and Locker’s model themselves. Critics of OHIP are sceptical of its ability to adequately represent OHQoL in the intended populations, let alone non-English-speaking cultures. For instance, it cannot factor in the positive or the many personal and environmental factors of oral health, including diet, economic priorities, personal expectations, health values and beliefs (Bakers, 2007; Brondani et al., 2007). Moreover, it remains unclear whether or not the ICIDH-based Locker’s model can accommodate cross-cultural differences in perceptions of disability and handicap (MacEntee et al., 1997). Therefore, it is safe to assume that culturally adapted versions of OHIP are fraught with similar limitations to the original English version.

Another limitation of Locker’s model is its sole focus on oral diseases. Consequently, the scope of Locker’s model is limited to the interactions between diseases and negative consequences, leaving no room for coping and adapting abilities or individual reactions to the same disease.
Despite these limitations and shortcomings, validating an existing scale required an appraisal in accordance with the original theoretical framework for which it was developed.

Consequently, my data did not include the representativeness of OHIP-14K or its degree of comprehensiveness in the item pool. Polit and Beck (2006) noted a similar limitation in their study, where they concluded that CVI ratings did not certify the inclusion of a comprehensive set of items to represent the construct of interest. Likewise, one SME commented that the OHIP-14K items did not adequately capture aesthetic aspect of oral health, which, according to the same SME’s opinion, was very significant to Korean older adults.

Meanwhile, there are alternative frameworks of oral health that could provide a more comprehensive and culturally universal foundation for OHQoL research in cross-cultural settings. ICF is an international classification system that holds the biopsychosocial view of health in which individual selves and social environments are shaped by the constant dynamics between them (Bury, 1982). Unlike its predecessor, ICF accounts for positive experiences and beliefs about health as well as the influence of subjective and environmental factors. For instance, ICF emphasizes the positive components of health, such as activity at the individual level and participation in society, and further states that personal factors (including gender, age, coping styles, social background, education, profession and lifestyle) interact with environmental factors (such as climate, cultural values, social attitude and structure) to affect bodily functions and structures/impairments, activity/limitations and participation/restriction (WHO, 2002). The interactions between various health dimensions are denoted by bidirectional arrows arranged in a non-linear fashion (Figure 6) and can be contrasted to the unidirectional progression presented by the ICIDH in Figure 2. This self-directed yet complex view of health in ICF is crucial for understanding QoL because it can account for QoL’s positive spectrum and the use of coping and adapting strategies for various oral conditions.
(Brondani & MacEntee, 2007), which help fight the stigma associated with disability (MacEntee, 2006).

Moreover, ICF domains are said to be culturally appropriate and universally applicable because the main principles of ICF include respecting different languages and cultures as well as conceptualizing functioning, health and disability on the basis of consensus from various perspectives, including health care, education, social policy, economics, insurance, and labour (Peterson, 2012).

Figure 6. The International Classification of Functioning, Disability and Health (ICF) (WHO, 2010b).

On the other hand, the major criticism of ICF stems from the fact that it is too complex and impractical to apply in daily practice (Üstün et al., 2010). It can be very difficult to operationalize for scale development (Ostir et al., 2006) and insensitive to cultural differences (Ingstad & Reynolds, 1995) and subjective dimensions (Ueda & Okawa, 2003).
Nonetheless, there are new, preeminent models of oral health that have emerged from the ICF classification. For example, the existential model of oral health, put forward by MacEntee (2006) (Figure 6) and later refined by Brondani, Bryant, and MacEntee (2007) (Figure 7), is designed to reflect the biopsychosocial aspects of oral health in various contexts associated with personal and socio-cultural factors.

The existential model expands on the language of the ICF and provides a more positive conceptualization of the various newly identified environmental and individual factors pertaining to oral health, including expectation, diet, economic priorities, health values and beliefs and influences of personal and social environments (Brondani et al., 2007). As can be seen in Figure 7, the key components of the refined model of oral health were empirically based and included a broader scope of experiences and oral health beliefs from older adults, with the limitation that its empirical basis was mostly Caucasians living independently in Vancouver, BC.
Figure 6. An existential model of oral health (MacEntee, 2006) [Reproduced with the permission of the editor of Gerodontology].
Figure 7. The atomic model of oral health (Brondani et al., 2007) [Reproduced with the permission of the editor of Gerodontology]
Chapter 7: Conclusion

The content validation method I used allows me to conclude that

- The OHIP-14K demonstrated limited evidence of content validity and cultural equivalency, and potential cross-cultural biases have been identified in its method, items, and construct representation.

- Further refinement of the instructions, response format, and items 5, 6, and 9 are necessary to improve the validity and equivalence of OHQoL measurement in Korean.

- The CVI technique is an effective tool for evaluating content validity and equivalency of a translated version of the OHIP for culturally specific and general considerations; detecting content biases in the early stages of cultural adaptation within budgetary realities to save cost, effort and time; and documenting the content validation process and quantification of CVIs and disagreement indices for future studies.
7.1 Study Implications

My study results suggest that current cultural adaptation and validation strategies for OHIP may not warrant adequate content validity or cultural equivalency for making valid and reliable cross-cultural comparisons of OHQoL. The main implication for cross-cultural measurement is the biased representation of the construct across cultures; therefore, caution is needed when comparing the OHIP scores cross-culturally. The differences in the OHIP score between cultural groups could be attributed to practical cultural differences in OHQoL, but also to construct-irrelevant variations, such as discrepancies in the method, items or construct.

Rigor is also needed for validating a translated version of a scale, much as it is used for developing a new scale, to ensure the compatibility and comparability of OHQoL measurement outcomes. Standardization of scale content and methods is critical to the validity of cross-cultural comparisons of psychometric constructs (King et al., 2011). The CVIs in conjunction with disagreement indices were useful tools for detecting and deriving solutions to various types of biases and for mediating disagreements between the experts. The expert suggestions and the CVI ratings could be used also to improve the content validity and equivalency of the OHIP-14K, as well as to refine inferences made from cross-cultural measurements of OHQoL.
7.2 Future Directions

It remains critical to consult lay Koreans to examine the relevance and utility of the OHIP-14K revisions suggested in this study and the implications of Locker’s theory of oral health for the target setting and purpose. Such social implications of the scores have yet to be considered, and further assessment of the refined version presented here should be undertaken through personal interviews or focus groups in a Korean population. In fact, WHO (2010a) specifically recommends conducting cognitive interviews with members of a target culture for making final decisions on words or phrases that conform better to their language. Using the cognitive interviewing method, the items that reflected high levels of disagreement could be appraised and likely resolved through consensus between lay respondents. Alternatively, focus groups can be sought to promote a moderated group discussion about the content validity of the study findings. The main advantage of focus groups is interactions among participants that allow more in-depth discussion of events (Casey & Kreuger, 2000). In the literature, focus groups have been used to content-validate Malay and Chinese versions of OHIP (Wong et al., 2002; Shiu et al., 2003; Saub et al., 2005) and other OHQoL studies (Brondani et al., 2007; Brondani & MacEntee, 2007). Focus groups with Korean older adults can provide lay perspectives on the content validity of OHIP-14K as there is a need to confirm whether lay respondents understand the items the way they were intended and to verify respondents’ subjective perceptions and understandings of their oral health conditions (Brondani & MacEntee, 2007).

Moreover, future studies should establish content validity and cultural equivalency of other language versions of OHIP-14 (or other OHQoL scales) and further explore the utility of CVIs and disagreement indices for enhancing cross-cultural OHQoL research paradigms, as there is a need for a continuous evaluation of the scale for the intended target populations in their
natural environment (Brondani & MacEntee, 2007). As can be deduced from the low level of translation equivalence between the English and Korean versions of OHIP-14, a comparison of two non-English versions might reveal similar problems, or perhaps even more so because current cultural adaptation and validation methods vary across the different-language versions of the OHIP. Hence, further inquiries into the equivalency of other language versions of OHIP would be necessary to ensure the validity of cross-cultural comparisons. Future content validation studies of OHQoL measures could also benefit from consulting dental hygienists or psychologists who can provide relevant information from different professional perspectives.

Now that some of the potential cross-cultural biases in the OHIP-14K have been identified by SMEs, empirical investigations can be directed at substantiating or refuting the impact of the detected biases on self-reporting carried out by non-expert respondents. Both the original and refined OHIP-14K can be administered to a sample of bilingual Koreans to confirm my content validity findings by comparing scores or factors. Factor analysis is commonly used to describe correlations with unobserved factors in data by examining equivalence in the factor loading of individual items and the equality of factorial relations (Byrne & Campbell, 1999). For instance, factor analysis has been applied to test the cross-cultural invariance of psychometric properties in a multi-dimensional measurement of social support across African American, non-Latino white, Chinese and Latino groups, with the authors of the study attaining supporting evidence for the presence of common factors, unidimensionality of items within each domain and invariance of each factors and its item set (Wong et al., 2010). Factor analysis is yet to be widely adopted in the validation of OHQoL scales in cross-cultural settings since it can be too costly and laborious. Moreover, without the knowledge of some of the biases I have presented, the factor analytic approach alone could not account for what may have caused the cross-cultural differences.
Final Comments

As an aspiring dentist and researcher, a parallel can be drawn between this learning and the very purpose of OHQoL measurement, which can be characterized as an effort to reconcile disagreement between patients’ perceptions and clinicians’ judgments of oral health. The self-report measures of OHQoL provide dental professionals with opportunities to listen to patients about health issues that extend far beyond what can be observed in clinics, and encourages them to work in partnership with their patients to protect their overall health. Although I have to yet enter a dental professional licensing body, I plan to put this holistic view of health into practice and use it to build a rapport with future patients.

I also came to realize that it is not easy to reach consensus even between experts – for example, dentists. Although information gathered from the panel of experts was more insightful and comprehensive than individual efforts, collective agreement was difficult to reach in the group decision-making process. I will always keep in mind that disagreement is an inevitable part of human interaction and that I need to skilfully mediate the conflict in order to successfully serve the best interests of patients.

Additionally, my graduate experience allowed me to develop creativity as well as interpersonal and problem-solving skills in the process of formulating and conducting my research. Following through the structured content-validation methods, I also learned to be aware of my own bias and stay open-minded in a scientific inquiry. During the literature review, I was overwhelmed by the vast amount of literature to be covered across different fields but came to appreciate the communality of multi-disciplinary studies and how seemingly different fields intersect, allowing for transfer of a method. Finally, I learned to accept the limitations of my own study and consider that the collection of all validity evidence as validation is not a
procedure but an ongoing process (Hubley & Zumbo, 1996). In this light, I shall bear in mind that validity claims of any test or measure should not be taken at face value and must undergo credible and careful scientific scrutiny.

Outside the classroom, I have experienced tremendous personal growth in terms of settling down in a new city and managing my academic life with my family. Vancouver offered ample opportunities to observe geriatric oral care in multicultural settings and to interact with colleagues and faculty members devoted to this field. Furthermore, caring for my own aging parents with chronic illnesses helped me choose a career path in geriatrics in line with my research as I enter an undergraduate dental educational degree. The overall graduate experience has broadened my perspective on dentistry as part of the health care fabric in Canada and has encouraged me to become a geriatric dentist.
References


and considerations in cross-cultural validation of generic health-related quality of life
instruments. In B. Spilker (Ed.) *Quality of life and Pharmacoeconomics in Clinical

Measuring Oral Health and Quality of Life (pp. 48-55). Chapel Hill, NC: Dental
Ecology, University of North Carolina.

number of items of the Oral Health Impact Profile on responsiveness, validity and
reliability in edentulous populations. *Community Dentistry & Oral Epidemiology, 36*,
12-20.

Validation of the Korean version of the oral health impact profile among the Korean
elderly. *Community Dentistry and Oral Epidemiology, 35*(1), 73-79.


Walker, R. J., & Kiyak, H. A. (2007). The impact of providing dental services to frail older adults: perceptions of elders in adult day health centers. *Special Care Dentistry*, 27, 139-143.


Appendices

Appendix A

A.1 Coding Search for Inclusion in Systematic Synthesis

<table>
<thead>
<tr>
<th>Data Base</th>
<th>Search Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEDLINE</td>
<td>Validation (90 citations)</td>
</tr>
<tr>
<td></td>
<td>((&quot;oral health&quot;[MeSH Terms] OR (&quot;oral&quot;[All Fields] AND &quot;health&quot;[All Fields]) OR &quot;oral health&quot;[All Fields]) AND Impact[All Fields] AND Profile[All Fields]) AND Validation[All Fields]</td>
</tr>
<tr>
<td></td>
<td>+ Translation (16 citations)</td>
</tr>
<tr>
<td></td>
<td>((&quot;oral health&quot;[MeSH Terms] OR (&quot;oral&quot;[All Fields] AND &quot;health&quot;[All Fields]) OR &quot;oral health&quot;[All Fields]) AND Impact[All Fields] AND Profile[All Fields]) AND Translation [All Fields]</td>
</tr>
<tr>
<td></td>
<td>+ Cultural adaptation (8 citations)</td>
</tr>
<tr>
<td></td>
<td>((&quot;oral health&quot;[MeSH Terms] OR (&quot;oral&quot;[All Fields] AND &quot;health&quot;[All Fields]) OR &quot;oral health&quot;[All Fields]) AND Impact[All Fields] AND Profile[All Fields]) AND Cultural adaptation[All Fields]</td>
</tr>
<tr>
<td>EMBASE</td>
<td>Validation (48 citations)</td>
</tr>
<tr>
<td></td>
<td>(Oral Health Impact Profile and Validation).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer]</td>
</tr>
<tr>
<td></td>
<td>+ Translation (13 citations)</td>
</tr>
<tr>
<td></td>
<td>(Oral Health Impact Profile and Translation).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer]</td>
</tr>
<tr>
<td></td>
<td>+ Cultural Adaptation (5 citations)</td>
</tr>
<tr>
<td></td>
<td>(Oral Health Impact Profile and Cultural Adaptation).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer]</td>
</tr>
</tbody>
</table>
B.1 English Version of Oral Health Impact Profile-14


HOW OFTEN have you had the problem during the last year? (circle your answer)

| 1. Have you had trouble pronouncing any words because of problems with your teeth, mouth or dentures? | VERY OFTEN | FAIRLY OFTEN | OCCASIONALLY | HARDLY EVER | NEVER | DON'T KNOW |
| 2. Have you felt that your sense of taste has worsened because of problems with your teeth, mouth or dentures? | VERY OFTEN | FAIRLY OFTEN | OCCASIONALLY | HARDLY EVER | NEVER | DON'T KNOW |
| 3. Have you had painful aching in your mouth? | VERY OFTEN | FAIRLY OFTEN | OCCASIONALLY | HARDLY EVER | NEVER | DON'T KNOW |
| 4. Have you found it uncomfortable to eat any foods because of problems with your teeth, mouth or dentures? | VERY OFTEN | FAIRLY OFTEN | OCCASIONALLY | HARDLY EVER | NEVER | DON'T KNOW |
| 5. Have you been self conscious because of your teeth, mouth or dentures? | VERY OFTEN | FAIRLY OFTEN | OCCASIONALLY | HARDLY EVER | NEVER | DON'T KNOW |
| 6. Have you felt tense because of problems with your teeth, mouth or dentures? | VERY OFTEN | FAIRLY OFTEN | OCCASIONALLY | HARDLY EVER | NEVER | DON'T KNOW |
| 7. Has your diet been unsatisfactory because of problems with your teeth, mouth or dentures? | VERY OFTEN | FAIRLY OFTEN | OCCASIONALLY | HARDLY EVER | NEVER | DON'T KNOW |
| 8. Have you had to interrupt meals because of problems with your teeth, mouth or dentures? | VERY OFTEN | FAIRLY OFTEN | OCCASIONALLY | HARDLY EVER | NEVER | DON'T KNOW |
| 9. Have you found it difficult to relax because of problems with your teeth, mouth or dentures? | VERY OFTEN | FAIRLY OFTEN | OCCASIONALLY | HARDLY EVER | NEVER | DON'T KNOW |
| 10. Have you been a bit embarrassed because of problems with your teeth, mouth or dentures? | VERY OFTEN | FAIRLY OFTEN | OCCASIONALLY | HARDLY EVER | NEVER | DON'T KNOW |
| 11. Have you been a bit irritable with other people because of problems with your teeth, mouth or | VERY OFTEN | FAIRLY OFTEN | OCCASIONALLY | HARDLY EVER | NEVER | DON'T KNOW |
12. Have you had difficulty doing your usual jobs because of problems with your teeth, mouth or dentures? 

<table>
<thead>
<tr>
<th></th>
<th>VERY OFTEN</th>
<th>FAIRLY OFTEN</th>
<th>OCCASIONALLY</th>
<th>HARDLY EVER</th>
<th>NEVER</th>
<th>DON'T KNOW</th>
</tr>
</thead>
</table>

13. Have you felt that life in general was less satisfying because of problems with your teeth, mouth or dentures? 

<table>
<thead>
<tr>
<th></th>
<th>VERY OFTEN</th>
<th>FAIRLY OFTEN</th>
<th>OCCASIONALLY</th>
<th>HARDLY EVER</th>
<th>NEVER</th>
<th>DON'T KNOW</th>
</tr>
</thead>
</table>

14. Have you been totally unable to function because of problems with your teeth, mouth or dentures? 

<table>
<thead>
<tr>
<th></th>
<th>VERY OFTEN</th>
<th>FAIRLY OFTEN</th>
<th>OCCASIONALLY</th>
<th>HARDLY EVER</th>
<th>NEVER</th>
<th>DON'T KNOW</th>
</tr>
</thead>
</table>
### B.2 Oral Health Impact Profile-14K (Bae et al., 2007)


<table>
<thead>
<tr>
<th>Last 3 months things due to oral problems you have had in the past week?</th>
<th>Very</th>
<th>Often</th>
<th>Occasionally</th>
<th>Rarely</th>
<th>Never</th>
</tr>
</thead>
</table>

※ Very: 1 week more than 2-3 times Usually: 1 week less than 2-3 times Occasionally: 1 week very rarely Rarely: 1 week or less Never: No experience

<p>| 1. Had difficulty eating when swallowing? | | | | | |
| 2. Did you notice any change in taste compared to before? | | | | | |
| 3. Did you notice any change in tooth sensation? | | | | | |
| 4. Did you notice any change in chewing sensation? | | | | | |
| 5. Did you notice any change in saliva? | | | | | |
| 6. Did you notice any change in oral breathing? | | | | | |
| 7. Did you notice any change in oral movement? | | | | | |
| 8. Did you notice any change in oral function? | | | | | |
| 9. Did you notice any change in oral appearance? | | | | | |
| 10. Did you notice any change in oral smell? | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>11. 다른 사람들에게 화를 잘 내게 되신 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. 평소 하시던 일을 하기가 어려웠던 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. 살아가는 것이 예전에 비해서 덜 만족스럽다고 느끼신 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. 정신적 신체적 사회적으로 전혀 제 루트를 할 수 없었던 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### B.3 Oral Health Impact Profile-14K (Kim & Min, 2009)

_Kim, G.U. & Min, K.J. (2009). The impact of oral health impact profile (OHIP) level on degree of patients' knowledge about dental hygiene. The Korean Academic Industrial Society, 10(4), 717-721._

#### Ⅲ. 구강건강에 관한 질문입니다. (해당란에 "V"해 주세요) OHIP—14

<table>
<thead>
<tr>
<th>최근 1년 동안 치아나 입안의 문제로 아래와 같은 느낌이나 문제가 발생한 적이 있었습니까?</th>
<th>매우 자주 있었다</th>
<th>자주있다</th>
<th>가끔</th>
<th>아주 가끔 있었다</th>
<th>전혀 없었다</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 입안의 문제로 발음 곤란을 느끼신 적이 있었습니까?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2. 입안의 문제로 맛을 느끼는데 어려운 적이 있었습니까?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3. 입안의 문제로 혀나 혀 밑, 림 입천장 등이 아픈 적이 있었습니까?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4. 입안의 문제로 음식물 먹기가 불편한 적이 있었습니까?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5. 입안의 문제로 타인을 만나기 꺼려 지신 적이 있었습니까?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6. 입안의 문제로 신경이 많이 쓰인 적이 있었습니까?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>7. 입안의 문제 때문에 식생활이 불만스러운 적이 있었습니까?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>8. 입안의 문제로 식사를 도중에 중단하신 적이 있었습니까?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>9. 입안의 문제로 식사 후 편히 쉬지 못한 적이 있었습니까?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>10. 입안의 문제 때문에 창피한 적이 있었습니까?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>11. 입안의 문제 때문에 타인에게 화를 내게 되신 적이 있었습니까?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>12. 입안의 문제 때문에 평소에 하시던 일을 하기가 어려운 적이 있었습니까?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>13. 입안의 문제 때문에 일상생활이 방해스럽지 못한 후 느껴진 적이 있었습니까?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
14. 입안의 문제 때문에 일상생활을 전혀 할 수 없었던 적이 있었습니까?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
</table>

### B.4 Oral Health Impact Profile-14K (Kim & Min, 2008)


```
<table>
<thead>
<tr>
<th>Question</th>
<th>Very Often</th>
<th>Often</th>
<th>Slightly</th>
<th>Not at all</th>
<th>Very Not at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Do you feel uncomfortable when you try to pronounce words?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Do you feel sensitive to taste changes compared to the past?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Do you feel uncomfortable when you brush your teeth?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Do you feel uncomfortable when you eat something hot or spicy?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Do you feel uncomfortable when you chew something?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Do you feel uncomfortable when you eat something cold or sticky?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Do you feel uncomfortable when you eat something sour or bitter?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Do you feel uncomfortable when you eat something crunchy or chewy?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Do you feel uncomfortable when you eat something soft or mushy?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Do you feel uncomfortable when you eat something crunchy or chewy?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Do you feel uncomfortable when you eat something crunchy or chewy?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Do you feel uncomfortable when you eat something crunchy or chewy?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
13. 입안의 문제 때문에 일상생활이 만족스럽지 못하다고 느껴친 적이 있습니까?

14. 입안의 문제 때문에 일상생활을 전혀 할 수 없었던 적이 있습니까?
Appendix C

C.1 Computation of Item-level $AD_M$ for Clarity and Cultural Equivalence Index

$AD_M$ for an item $j$ is estimated as follows:

$$AD_{M(j)} = \frac{\sum_{k=1}^{N} |x_{jk} - \bar{x}_j|}{N},$$

where $N$ is the number of judges or observations for an item $j$, $x_{jk}$ is the $k$th judge’s rating on item $j$, and $\bar{x}_j$ is the mean of the judges’ scores on item $j$. (Burke & Dunlap, 2002)

C.2 Computation of $AD_M$ for Clarity and Equivalence Index

$$AD_{M(J)} = \frac{\sum_{j=1}^{J} AD_{M(j)}}{J}.$$

or simply averaged $AD_{M(j)}$ (Burke & Dunlap, 2002).
C.3 Computation of Multirater $K_{free}$ for Relevance index items

\[
K_{free} = \frac{1}{Nn(n-1)} \left( \sum_{i=1}^{N} \sum_{j=1}^{k} n_{ij} - Nn \right) - \left[ \frac{1}{k} \right]
\]

\[
1 - \left[ \frac{1}{k} \right]
\]

$N$ is the number of cases, $n$ is the number of raters, and $k$ is the number of rating categories. (Randolph, 2005)

C.4 Interpretation of Multirater $K_{free}$ Statistics

<table>
<thead>
<tr>
<th>Kappa</th>
<th>Level of agreement (Landis &amp; Koch, 1977)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&gt; 0$</td>
<td>Equal to chance and poor agreement</td>
</tr>
<tr>
<td>0.00 – 0.20</td>
<td>Slight agreement</td>
</tr>
<tr>
<td>0.21 – 0.40</td>
<td>Fair agreement</td>
</tr>
<tr>
<td>0.41 – 0.60</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>0.61 – 0.80</td>
<td>Substantial agreement</td>
</tr>
<tr>
<td>0.81 – 1.00</td>
<td>Almost perfect agreement</td>
</tr>
</tbody>
</table>
Appendix D

D.1 Informed Consent

Consent Form (Content Experts)

RESEARCH TITLE: Qualitative Validation of Korean version of OH-QoL Measures:

Content Validation with Korean Content Experts

Principal: Jaesung Seo, HBSc

Investigator: MSc Candidate at the Faculty of Dentistry, UBC

Phone: 604-263- XXXX

Research Supervisors: Dr. Mario Brondani, DDS, MSc, PhD

Dr. Michael MacEntee, LDS, Dlp. Prosth, FRCD, Ph.D.

The following information has been provided to you for making an informed decision about your participation in this study.

PURPOSE

The purpose of this study is to conduct a content validation study of the Korean version of oral health impact profile (OHIP), designed to measure your oral health-related quality of life (OH-QoL). It is being conducted as part of dissertation for Masters in Science degree.
POTENTIAL BENEFITS

The findings of this research can help delineate common problems in cultural adaptation of OHQoL measures and propose the use of structured content validation methods to improve validity of cross-cultural measurement and comparison of OHQoL.

STUDY PROCEDURE

You are invited to self-administer the content validation questionnaire electronically on the attached PDF form. The content validation questionnaire is designed to collect information about the content validity of two Korean versions of OHIP. You will be asked to rate each element of the indicators for clarity, relevance, and cultural equivalence. Upon completion of the questionnaire, you can submit the PDF form through FTP server by clicking on the submit button or E-mail the saved PDF form to _______@interchange.ubc.ca.

CONFIDENTIALITY

Please note that you are under no obligation to participate in this study. If you choose to participate, you can withdraw or ignore questions at any time without consequences. Any information obtained in this study and that can be identified with you will remain confidential and will be disclosed only with your permission. Communication in scientific reports will identify the program and individual participants only by a code known to the investigators. Information you give will be stored confidentially, and under no circumstances will it be revealed to other faculty members without your expressed wish and instructions. CDs of audio-recordings and transcripts of interviews will be stored in a locked filing cabinet at UBC and will be destroyed after five years.

KNOWN RISKS
There are no known risks associated with participating in this research. Data collection will take place only after you are fully aware of the study, and after your informed consent is obtained.

**REMUNERATION/COMPENSATION**

You will be remunerated $50 for your participation.

**CONTACT FOR INFORMATION ABOUT THE STUDY**

If you have any questions or would like further information with respect this study, you may contact:

Jaesung Seo

Phone: 604-XXX-XXXX

E-mail: _________@interchange@ubc.ca

**CONTACT FOR CONCERNS ABOUT THE RIGHTS OF RESEARCH PARTICIPANTS**

If you have any concerns about your treatment or rights as a research participant, you may contact the Research Subject Information Line in the UBC Office of Research Services at 604-XXX-XXXX or ______@ors.ubc.ca.

**CONSENT**

Your participation in this study is entirely voluntary, and you may refuse to participate or withdraw from the study at any time without consequence. You do not have to answer all questions on the questionnaires.

Your signature below indicates that you have received a copy of this consent form for your own records.
Your signature indicates that you consent to participate in this study. You are willing to have your interview audio-recorded and give permission for the principal investigator to use the information you are providing as part of a larger study focused on the same issue.

X

{Participant’s signature}  {Date}

X

{Printed name of the participant signing above}

{Principal Investigator’s signature}  {Date}
D.2 Cover Letter and Content Validation Questionnaire

To whom it may concern:

As part of my MSc program at the UBC Faculty of Dentistry, I am undertaking content validation of the Korean versions of oral health impact profile-14 (OHIP-14). As the demographic population in Korea is rapidly aging, the need for reliable and valid measurement of oral health-related quality of life (OH-QoL) has greater significance for setting goals in oral health and evaluating the degree to which those goals are met.

You are invited to participate in this content validation study as a subject matter expert because of your expertise and experience in clinical dentistry. Your contribution to this study is valuable in improving not only the Korean versions of OH-QoL measures but also the current standards for translation and cultural adaptation of the measures into different languages for accurate cross-cultural measurement and comparison.

OHIP, the most widely used OH-QoL measure worldwide, is one of few instruments translated to and validated in Korean. OHIP is based on Locker's model of oral health derived from WHO's international classification of impairments, disabilities, and handicaps. OHIP consists of 7 dimensions including functional limitations, physical pain, psychological discomfort, physical disability, psychological disability, social disability, and handicap. OHIP is assessed with a five point Likert scale, with 0 representing never and 4 representing very often.

You will be asked to evaluate the Korean instructions, response formats, and items on clarity, relevance, and equivalence with the English version.

Clarity
Please rate the clarity of instructions, response scale, and items for more accurate data collection. Examples of threats to clarity include double-barreled questions, dental jargon, cognitive-challenging questions, and culturally inappropriate expressions.

Relevance
Please rate the relevance of each item to the dimension or the construct. You should unbiased measurement of OH-QoL at the level of methods, items, and constructs in comparison to the original English version. Equivalence can be examined in terms of equivalence between the Korean and English versions in meaning (semantic equivalence), but also in idiomatic expression (idiomatic and colloquial equivalence), relevant situations (experiential equivalence), and concepts (conceptual equivalence). Bias, on the other hand, is a factor contributing to score differences other than cultural differences. Please report any potential bias in the text box provided.

If you wish to save your progress and continue later, please click SAVE under FILE or press CTRL + S.

Should you have any questions regarding this questionnaire, please contact Jan at jmr@ubc.ca or Dr. Mario Brondani (MSc Supervisor) at 604-822-2 / hrc@change.ubc.ca.

Content Validation of the Korean Versions of OH-QoL Indicators
### CLARITY OF QUESTIONNAIRE ELEMENTS

Please click on your rating of each Indicator Element on clarity:

<table>
<thead>
<tr>
<th>Indicator Element</th>
<th>Not at all Clear</th>
<th>Somewhat Clear</th>
<th>Mostly Clear</th>
<th>Very Clear</th>
<th>Suggested additions, deletions, and modifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructions</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>中断3개월 동안 치아이나 잇몸에 어파의 경험을 얼마나 자주 겪으셨습니까?</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Response Formats</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>머무, 차주, 가끔, 거의, 전혀</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Response Format Labels</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>※ 머무: 1주일에 2-3회 이상 자주: 1주일에 1회 정도 가끔: 한 달에 2-3회 정도 거의: 한 달에 1회 이하 전혀: 전혀</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Q 1. 일일이 치아이 보관해 풀린 염이 있습니까?</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Q. 2 왕을 느끼는 감각이 예전보다 더해졌다고 느끼신 건이 있습니까?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Q. 3 아니 해밀, 투입감정 등이 아픈 건이 있습니까?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Q. 4 아프거나 기분이 떨어질 때면 일주일 동안 히가 불편한 건이 있습니까?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Q. 5 즐거움이 다른 사람들을 만나기로 거리조심의 점이 있습니까?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Q. 6 신경이 많이 흔한 점이 있습니까?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Q. 7 심장이 불안스러운 점이 있습니까?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Q. 8 심사람도 조금은 증상이 있습니까?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Q. 9</td>
<td>편안하게 미지를 못하신 적이 있습니까?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q. 10</th>
<th>난처하거나 멍석스러웠던 적이 있습니까?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>○</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q. 11</th>
<th>다른 사람들에게 화를 갈 내게 되신 적이 있습니까?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>○</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q. 12</th>
<th>절소 하시던 일을 하기로 이레했던 적이 있습니까?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>○</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q. 13</th>
<th>살아가는 것이 어색해 비헤시 및 민족스럽다고 느끼신 적이 있습니까?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>○</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q. 14</th>
<th>절소적 질체적 사회적으로 정직하게 도와할 수 없었던 적이 있습니까?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>○</td>
</tr>
</tbody>
</table>
WHO’s framework of disability known as International Classification of Impairments, Disabilities, and Handicaps (ICIDH) (1980). ICIDH explains consequences of disease in three dimensions:

- Impairment: organ system level loss of structure or function
- Disability: person level loss of the ability to function in ways considered normal for a human being
- Handicap: societal level disadvantage for the person created by the intersection of the impairment or disability with the environment and that person’s roles.
### RELEVANCE OF QUESTIONNAIRE ITEMS TO DIMENSION

Please choose the dimension that each item best represents from the drop box.

<table>
<thead>
<tr>
<th>Items</th>
<th>Dimension</th>
<th>Suggested additions, deletions, and modifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q. 1. 복음이 잘 인지되어 부림했었던 적이 있으신가요?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q. 2. 많은 느끼는 감각이 예전보다 나빠졌다고 느껴신 적이 있습니까?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q. 3. 하니 형질, 발질환 등이 아픈 적이 있습니까?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q. 4. 아프거나 기분전환 압력의 문제 때문에 출석, 연기가 불편한 적이 있습니까?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q. 5. 할게해서 다른 사람을 만나기에 관리하는 적이 있습니까?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q. 6. 실패하면 많이 쓰인 적이 있습니까?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q. 7. 실패하면 불안스러운 적이 있습니까?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q. 8. 실수를 두중에 중단하신 적이 있습니까?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q. 9. 전달하게 되지 못하신 적이 있습니까?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q. 10 난처하거나 말할수없던 적이 있습니까?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q. 11 다른 사람들에게 화를 가리게 되신 적이 있습니까?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q. 12 핑소 하시던 일을 하다가 어려웠던 적이 있습니까?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q. 13 살아가는 것이 어렵게 되면서 일에 만족스럽다고 느껴진 적이 있습니까?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q. 14 질병의 심각성의 상황으로 전력 제한을 받은 적이 있습니까?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**EQUIVALENCE BETWEEN KOREAN AND ENGLISH VERSIONS**

Please evaluate the equivalence between the Korean that each item best represents from the drop box.

<table>
<thead>
<tr>
<th>Indicator Element</th>
<th>Not at all Equivalent</th>
<th>Somewhat Equivalent</th>
<th>Mostly Equivalent</th>
<th>Equivalent</th>
<th>Suggested additions, deletions, and modifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>〇 〇 〇 〇</td>
</tr>
<tr>
<td>Response Formats</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>〇 〇 〇 〇</td>
</tr>
<tr>
<td>Response Format Labels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Content Validation of the Korean Versions of OH-Qol Indicators
| Q. 매주 1 주일에 2-3 회 이상 귀주: 1 주일에 1 회 | ○ | ○ | ○ | ○ |
| Q. 2 | ○ | ○ | ○ | ○ |
| Q. 3 | ○ | ○ | ○ | ○ |
| Q. 4 | ○ | ○ | ○ | ○ |
| Q. 5 | ○ | ○ | ○ | ○ |
| Q. 6 | ○ | ○ | ○ | ○ |

Not specified in the English version.

Q. 1. Have you had trouble **pronouncing any words** because of problems with your teeth, mouth or dentures?

Q. 2. Have you felt that your **sense of taste** has worsened because of problems with your teeth, mouth or dentures?

Q. 3. Have you had **painful aching** in your mouth?

Q. 4. Have you found it **uncomfortable to eat any foods** because of problems with your teeth, mouth or dentures?

Q. 5. Have you been **self conscious** because of your teeth, mouth or dentures?

Q. 6. Have you felt **tense** because of problems with your...
<table>
<thead>
<tr>
<th>Q.</th>
<th>Question</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Has your diet been unsatisfactory because of problems with your teeth, mouth or dentures?</td>
<td>○ ○ ○ ○ ○ ○</td>
</tr>
<tr>
<td>8</td>
<td>Have you had to interrupt meals because of problems with your teeth, mouth or dentures?</td>
<td>○ ○ ○ ○ ○ ○</td>
</tr>
<tr>
<td>9</td>
<td>Have you found it difficult to relax because of problems with your teeth, mouth or dentures?</td>
<td>○ ○ ○ ○ ○ ○</td>
</tr>
<tr>
<td>10</td>
<td>Have you been a bit embarrassed because of problems with your teeth, mouth or dentures?</td>
<td>○ ○ ○ ○ ○ ○</td>
</tr>
<tr>
<td>11</td>
<td>Have you been a bit irritable with other people because of problems with your teeth, mouth or dentures?</td>
<td>○ ○ ○ ○ ○ ○</td>
</tr>
<tr>
<td>12</td>
<td>Have you had difficulty doing your usual jobs</td>
<td>○ ○ ○ ○ ○ ○</td>
</tr>
</tbody>
</table>
### ADDITIONAL FEEDBACK

What is your overall opinion on the clarity of the Korean version of OHIP?

What is your overall opinion on the relevance of the Korean version of OHIP to the construct of Oral Health Related Quality of Life?

What is your overall opinion on the equivalence between the Korean version of OHIP and the original English version?
Additional Comments:

Thank you very much for taking the time to complete this content validation form.

Please Return Completed Form via E-Mail/FAX/Mail to: