



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

조 영 일 교수 지도

박사학위 청구논문

건강관련 삶의 질 측정도구에 대한  
PLS-SEM의 적용

2017

성신여자대학교 대학원

심 리 학 과

김 혜 영

건강관련 삶의 질 측정도구에 대한  
PLS-SEM의 적용

조영일 교수 지도

이 논문을 박사학위논문으로 제출함

2017년 4월

성신여자대학교 대학원

심리학과

김혜영

# 인 준 서

김혜영의 박사학위 논문으로 인준함

2017년 4월

심사위원장 \_\_\_\_\_ (인)

심사위원 \_\_\_\_\_ (인)

심사위원 \_\_\_\_\_ (인)

심사위원 \_\_\_\_\_ (인)

심사위원 \_\_\_\_\_ (인)

성신여자대학교 대학원

## 논문개요

사람의 건강을 평가함에 있어 전통적으로 사용하였던 객관적 임상검사 결과와 더불어 환자 본인의 주관적인 보고를 포함한 포괄적인 결과물 평가가 일반화함에 따라 환자의 관점에서 본인이 안녕(well-being)하다고 주관적으로 느끼는 정도를 평가하는 건강관련 삶의 질(Health-Related Quality of Life, HRQoL)의 중요성이 부각되고 있다.

HRQoL 도구는 다차원적인(multidimensional) 도구로서 각각의 하부개념들은 반영적 또는 형성적 측정구조를 가질 수 있고 그러한 하부개념들을 종합하는 HRQoL은 하부개념들과 형성적 관계를 가지는 것으로 보인다. 그러나 기존 연구에서 HRQoL 도구들은 심리측정적 전통을 따라 반영적인 구조와 관계로만 설정되어 연구되고 있다.

본 연구에서는 HRQoL 도구로 대표적인 Short Form-36(SF-36), EuroQoL-5D(EQ-5D), Oral Health Impacts Profile for Children(COHIP) 및 Oral Impacts on Daily Performances for Children(COIDP)의 측정구조를 심층적으로 고찰하여 내용과 부합하는 적절한 구조방정식 모형으로 도출한 후 실증적인 자료를 이용하여 Partial Least Square Structural Equation Model (PLS-SEM)을 적용하여 분석하였다.

첫째, 본 연구에서는 다양한 HRQoL 측정도구를 내용적으로 검토하고 각 차원의 측정구조 및 차원들과 상위 개념과의 관계가 반영적인지 형성적인지를 고찰함으로써 HRQoL 측정도구 각각에 적합한 구조모형을 수립하여 이를 실증 분석의 기초 모형으로 사용하였다.

고찰의 결과, SF-36의 일차 구성개념들은 반영적 측정구조를 가지며, 일차 구성개념과 이차 구성개념은 형성적 관계를 가지는 구조모형으로 규정하였다. COHIP의 일차 구성개념들은 반영적 측정구조와 형성적 측정구조가 혼재

된 것으로 보았고, 일차 구성개념과 이차 구성개념은 형성적 관계를 가지는 구조모형으로 규정하였다. EQ-5D와 COIDP는 형성적 측정모형으로 규정하였다.

둘째, PLS-SEM은 형성적 측정모형의 원리에 의하여 복합지수를 기반으로 추정하므로 반영적 특성을 가진 자료에 PLS-SEM을 적용하였을 때 발생하는 추정 오류의 정도를 평가하고자 하였다. 가상 데이터 생성 및 분석 결과, 요인별 지표변수의 수가 3개이면서 요인계수가 0.6 이하인 모든 경우와 요인계수가 0.4로 낮으면서 동시에 요인별 지표변수의 수가 6개 이하인 경우 심각한 추정 오류를 나타내었다. 분석 대상인 HRQoL 도구 중 반영적 측정구조를 포함하는 SF-36과 COHIP의 자료들은 요인계수가 비교적 크고 요인 당 지표변수의 수도 커서 예상되는 오류수준이 크지 않을 것으로 판단하였다.

셋째, PLS-SEM을 이용하여 HRQoL 측정도구에 대한 실증분석을 실시한 결과 모형이 잘 적합되었다. 상기한 HRQoL 도구에 대한 PLS-SEM 적용 결과 반영적 측정모형은 내적일관성, 지표변수 신뢰성, 수렴타당성 및 판별 타당성이 대체로 충족되었고 형성적 측정모형을 가지는 차원 역시 다중공선성 및 외부 가중치 타당성 면에서 적절하였다. 일차 구성개념들과 이차구성개념이 형성적 관계로 정의된 이차 계층 구조를 포함한 전체 구조방정식에도 PLS-SEM이 잘 적용되었다.

넷째, 여러 차원들을 다양한 방법으로 통합하여 HRQoL로 요약점수로 표현하여 비교한 결과, SF-36 및 COHIP 도구의 경우 PLS-SEM모형에서 산출된 차원 간을 통합하는 가중치를 이용한 경우 일반적으로 요약점수로 사용되고 있는 차원점수 평균이나 전체 문항점수 평균과의 상관 정도가 매우 높은 것으로 나타났다. PLS-SEM 적용 결과 산출된 가중치로 계산한 요약점수는 EQ-5D 도구의 선호도 기반 점수와 비교하였을 때에도 상당한 정도로 유사하였다. 따라서 본 연구의 결과에 따르면 사용자의 편리성 면에서 유리한 합산 점수나 평균 점수를 구성개념의 요약점수로 사용하는 것이 대체로 적절하다고

사료되었다.

본 연구에서는 대표적인 HRQoL 도구들을 분석함에 있어서 전체 측정도구의 구조를 아우르는 통계적 모형을 적용하였다. 통계적 모형 적용의 장점은 다음과 같다.

첫째, 측정모형의 타당성 평가가 가능하다. 모형 적용 후 분석 결과에 대하여 다양한 신뢰도 및 타당도 검증방법을 적용함으로써 전체 측정구조의 적합도, 개별 변수의 신뢰도 및 구성개념의 타당도 등을 평가할 수 있다. 둘째, 다양한 표본에 대한 분석 결과를 비교하거나 평가할 수 있고 교차타당도 등의 검증을 통하여 측정모형의 일반화가 가능하다는 장점이 있다. 교차타당화를 시도하고 상이한 분석 결과에 대하여 차이의 원인을 규명하는 과정은 해당 도구의 측정모형에 대한 일반화를 위한 노력의 출발점이 될 수 있다.

본 연구에서는 HRQoL 도구의 측정모형에 PLS-SEM 모형을 적용하였는데 구성개념에 대한 측정모형의 관점에서 볼 때 몇 가지 의문점이 제기되었다. 첫째, 모형 식별을 위하여 함께 포함된 다른 변수로 인하여 추정치의 큰 변화가 나타난다는 점이다. 둘째, 반영적 측정모형의 경우, 수렴타당도 기준이 평균추출분산(average variance extracted) 0.5인데 이는 전체 지표변수 분산의 50% 만을 공통성으로 한다. 그럼에도 불구하고 지표변수의 전체 변동을 포함하는 복합지수(composite)를 구성개념 값으로 이용하는 것은 상호 모순적일 수 있다.

셋째, 형성적 모형의 수렴타당성을 평가하기 위하여 중복 분석이 시행되었다. 그러나 현실적으로 서로 다른 방법으로 조사한 유사 구성개념간의 구조계수가 0.8을 상회하기는 쉽지 않으므로 중복 분석의 실효성에 의문을 제기할 수 있다. 넷째, 계층구조인 도구에서 형성적 관계의 이차 구성개념을 가지는 경우 모형의 결정계수는 대부분 1이 산출되어 모형의 예측력이나 적합도를 평가하는 데 활용할 수 없었다. 다섯째, 특수 분산 간 상관을 모형에 반영하고 있는

CB-SEM과 달리 PLS-SEM에서는 이러한 특수 분산 간 상관의 설정이 가능하지 않다.

이 연구주제와 관련하여 필요하다고 여겨지는 후속 연구는 첫째, 이 연구의 방법론을 보건학 분야의 다양한 구성개념으로 확대 적용하여 그들의 구조를 이론적으로 고찰하고 또한 경험적 데이터 분석을 실시하여 각 구성개념들의 올바른 구조 관계를 연구할 필요성이 있다. 둘째, 이차 계층구조를 포함하여 다양한 조건과 구조를 설정하고 모의실험을 실시하는 연구의 필요성이 있다. 조건에 맞는 데이터를 생성하여 CB-SEM 및 PLS-SEM 적용 결과를 비교하고 그 수행을 평가하는 연구가 필요하다. 셋째, 이 연구에서는 반영적 및 형성적 측정 모형의 구별이 이론적 내용적인 면에서 한정되어 다소 주관적인 판단에 의하여 결정할 우려가 있으므로 판단 기준을 객관화시킬 수 있는 방법론의 개발이 요망된다. 넷째, PLS-SEM 전체 모형의 적합도 평가 방법을 정교화 하는 후속연구가 필요하다고 사료된다.

결론적으로 본 연구의 의의는 전통적으로 반영적 구조에만 의거하여 측정구조를 분석하던 여러 HRQoL 측정도구에 대하여 형성적 및 반영적 관계를 적절하게 규정하고 올바르게 정의된 측정모형에 PLS-SEM을 적용하여 경험적으로 구현하였으며 구현된 모형의 결과를 기반으로 다차원적 HRQoL 측정도구의 차원을 통합하는 방법을 제안하였다는 점에 있다.

**주요어:** 건강관련 삶의 질(HRQoL), Partial Least Square 구조방정식(PLS-SEM), 공분산 구조방정식, 형성적 측정모형, 복합지표변수.

# 목 차

I . 서론 .....	1
1. 연구의 필요성 및 목적 .....	1
II . 이론적 배경 .....	7
1. 건강관련 삶의 질 .....	7
1.1 건강관련 삶의 질 .....	7
1.2 임상연구의 결과변수로서의 건강관련 삶의 질 .....	8
1.3 건강관련 삶의 질 측정도구에 대한 두 접근법 .....	01
1.4 건강관련 삶의 질 도구의 종류 .....	5
1.5 질보정 수명을 이용한 효과 평가 .....	2
1.6 정신건강과 QALY .....	23
1.7 측정 관점에서의 고찰 .....	4
2. 사회과학에서의 측정이론의 종류 .....	8
2.1 일관행위론 .....	8
2.2 수량산출론 .....	9
2.3 현실표상론 .....	1
3. 자연과학과 사회과학에서의 측정오차 .....	23
3.1 자연과학에서의 측정오차 .....	2
3.2 사회과학에서의 측정오차 .....	4
4. 구성개념과 측정모형의 구분: 형성적 및 반영적 .....	63
4.1 구성개념의 정의 .....	8
4.2 반영적 구성개념과 반영적 측정모형 .....	63
4.3 형성적 구성개념과 형성적 측정모형 .....	83
4.4 반영적 및 형성적 지표변수의 예 .....	24
4.5 반영적 측정모형과 형성적 측정모형의 차이점 .....	44

4.6 측정모형의 식별 오류 현황 .....	74
5. 반영적 및 형성적 측정모형의 통계적 추정 .....	94
5.1 CB-SEM의 통계적 추정 .....	94
5.2 PLS-SEM의 통계적 추정 .....	15
5.3 일반화 구조성분분석법 .....	3
6. 공통요인 접근법과 주성분 접근법의 수리적 비교 .....	45
6.1 추정 과정의 수리적 비교 .....	4
6.2 추정 결과의 수리적 비교 .....	6
7. PLS-SEM의 통계적 추정과 분석 결과 평가 .....	85
7.1 PLS-SEM 모형 산출 과정 .....	85
7.2 PLS-SEM 모형 추정 모수 .....	85
7.3 반영적 측정모형의 평가 .....	9
7.4 형성적 측정모형의 평가 .....	8
7.5 PLS-SEM 분석 결과 .....	66
7.6 2차 계층 구조의 추정과 평가 .....	76
III. 연구 목적 .....	72
IV. 연구 방법 .....	73
1. 다양한 HRQoL 측정도구의 특성 고찰 .....	73
2. 데이터 모의실험 연구 .....	2
3. 건강관련 삶의 질 측정 자료에 대한 PLS-SEM 분석 .....	47
4. 현행 방법과 PLS-SEM적용에 의한 통합점수의 비교 .....	57
V. 연구 결과 .....	76
1. 건강관련 삶의 질 측정도구 고찰 .....	67
1.1 36-item Short Form Health Survey(SF-36) .....	67

1.2 EuroQoL-5D(EQ-5D) .....	9
1.3 어린이 구강건강영향 프로파일 .....	79
1.4 어린이 일상활동 구강영향 .....	12
2. 반영적 측정모형 기반 데이터 생성 및 분석(모의실험) .....	51
2.1 반영적 측정모형 기반 데이터 생성 및 분석 .....	6
2.2 분석결과의 평가 .....	107
2.3 SF-36과 COHIP 경험적 자료와의 비교 .....	17
3. 건강 관련 삶의 질 측정 자료에 대한 PLS-SEM 분석 .....	91
3.1 반영적-형성적 측정구조: SF-36 .....	9
3.2 형성적 측정구조: EQ-5D .....	126
3.3 혼합적 측정구조: COHIP .....	130
3.4 형성적 측정구조: COIDP .....	136
3.5 COHIP 모형에 대한 교차타당화 .....	143
4. 기존 방법과 모형에 의한 구성개념 점수 비교 .....	146
4.1 SF-36 .....	144
4.2 COHIP .....	147
4.3 EQ-5D .....	148
VI. 논의 .....	151
1. 결과해석 및 의의 .....	151
2. 후속연구 제안 .....	158
3. 종합논의 및 결론 .....	160

참고문헌

ABSTRACT

부록

## 표 목차

표 1. 다양한 건강관련 삶의 질 측정도구 .....	51
표 2. 측정모형 선택에 대한 기준과 판단 .....	44
표 3. PLS-SEM모형 추정과 관련한 평가 사항 .....	56
표 4. SF-36 삶의 질 도구의 8개 차원 과 점수의 해석 .....	97
표 5. SF-36의 문항, 점수체계 및 차원 .....	08
표 6. SF-36의 주성분 분석 결과로 나타난 신체 및 정신주성분 .....	38
표 7. Short Form-6D .....	8
표 8. 오늘의 건강상태를 질문하는 EQ-5D 도구 .....	29
표 9. 어린이 구강건강영향 프로파일(COHIP)의 요인 및 항목 .....	89
표 10. 어린이 일상활동 구강영향(COIDP) 설문 .....	0
표 11. 반영적 자료의 주성분 분석 결과 평가 .....	15
표 12. SF-36 예시 자료의 요인계수의 크기 및 지표변수 수 .....	71
표 13. COHIP 예시 자료의 요인계수의 크기 및 지표변수 수 .....	8
표 14. SF-36 에 대한 반복지표변수 모형의 구조계수(n=188) .....	121
표 15. SF-36의 PLS-SEM 결과 평가 및 CB-SEM 결과와의 비교 .....	221
표 16. SF-36모형의 판별타당도(Fornell-Larker 기준) .....	421
표 17. SF-36 2차 계층모형의 PLS-SEM 추정 결과 평가 .....	421
표 18. EQ-5D의 형성적 측정모형의 질 평가 .....	17
표 19. EQ-5D 모형과 EQ-VAS .....	17
표 20. EQ-5D 형성적 측정모형 및 구조모형의 평가 .....	8
표 21. COHIP의 형성적 구조계수(반복지표변수 모형, Mode B) .....	3
표 22. COHIP의 형성적 1차 구조 'Oral Health'에 대한 질 평가 .....	12
표 23. COHIP의 반영적 1차 구조의 질 평가 기준 .....	13
표 24. COHIP 모형의 판별타당도(Fornell-Larker 기준) .....	4
표 25. COHIP 2차 계층모형의 PLS-SEM 추정과 관련한 평가 .....	6

표 26. COIDP 지표변수 간의 비모수적 상관계수 .....	17
표 27. COIDP의 형성적 측정모형의 질 평가 .....	19
표 28. COIDP의 형성적 측정모형과 COHIP .....	139
표 29. COIDP의 형성적 측정모형 및 구조모형의 평가 .....	9
표 30. COIDP의 반영적 1차 구조의 질 평가 기준 .....	11
표 31. 반영적 COIDP 모형과 COHIP .....	141
표 32. 반영적 COIDP 측정모형과 구조모형 평가 .....	12
표 33. COIDP 모형화에 사용된 COHIP 측정모형의 가중치 비교 .....	18
표 34. COHIP의 형성적 구조계수 비교 .....	143
표 35. COHIP의 형성적 1차 구조 'Oral Health'에 대한 요인계수 비교 .....	14
표 36. COHIP의 반영적 1차 구조의 요인계수 비교 .....	14
표 37. SF-36 차원, 문항 평균 및 가중치에 의한 통합 .....	61
표 38. 여러 방법에 의한 SF-36 통합점수의 피어슨 상관계수 .....	61
표 39. COHIP 차원, 문항 평균 및 가중치에 의한 통합 .....	17
표 40. 여러 방법에 의한 COHIP 통합 점수의 피어슨 상관계수 .....	1
표 41. 선호도기반 및 PLS-SEM 추정에 의한 EQ-5D의 점수 .....	81
표 42. EQ-5D의 수준별 점수 간의 스피어만 상관계수 .....	9

## 그림 목차

그림 1. Pubmed 검색 결과 나타난 ‘건강관련 삶의 질’ 관련 문헌의 추이 .....	8
그림 2. 건강관련 삶의 질 모형에서의 환자 결과물 측정 간의 관계 .....	0 1
그림 3. 시간교환법(Time Trade-Off, TTO) .....	4 1
그림 4. Pubmed 검색 결과 나타난 ‘SF-36’ 관련 문헌의 추이 .....	6 1
그림 5. Pubmed 검색 결과 나타난 ‘EQ-5D’ 관련 문헌의 추이 .....	91
그림 6. 반영적 측정모형 .....	8
그림 7. 형성적 측정모형의 두 종류 .....	04
그림 8. Multiple Indicators - multiple causes (MIMIC) 모형 .....	5
그림 9. 4 가지의 가능한 이차 측정모형 구조 .....	86
그림 10. PLS-SEM의 2차 계층모형의 추정을 위한 반복 지표변수 모형 .....	0· 7
그림 11. PLS-SEM의 2차 계층모형의 추정을 위한 Hybrid 모형 .....	0 7
그림 12. 기존 연구에 나타난 반영적 3차 SF-36 구조모형 .....	5· 8
그림 13. 한국어판 SF-36을 이용하여 추정된 반영적 SF-36 구조모형 .....	7· 8
그림 14. 한국어판 SF-36을 이용하여 추정된 3차 반영적 SF-36 구조모형 .....	8· 8
그림 15. 본 연구에서 가정하는 형성적 2차 구조를 가지는 SF-36 구조모형 .....	9· 8
그림 16. EQ-5D의 개념적 모형 .....	9
그림 17. EQ Visual Analogue Scale (EQ-VAS) .....	49
그림 18. 구강건강영향 프로파일(COHIP)의 공분산 기반 2차 계층모형 .....	9 9
그림 19. 본 연구의 구강건강영향 프로파일(COHIP) 분석 모형 .....	0
그림 20. 본 연구의 일상활동 구강건강영향(COIDP) 분석 모형 .....	0
그림 21. 요인계수 참값과 추정 평균의 차이 .....	D
그림 22. 요인계수 추정치의 표준오차 .....	D
그림 23. 요인계수 추정치의 평균 평균제곱오차(MSE) .....	41
그림 24. SF-36 측정도구의 PLS-SEM 분석 모형과 분석 결과 .....	021

그림 25. EQ-5D 도구의 형성적 측정모형 PLS-SEM 분석 .....	㉑
그림 26. EQ-5D 도구의 형성적 측정모형 주성분 분석 .....	㉒
그림 27. COHIP 도구의 혼합적 측정모형 및 계층적 구조모형 분석 .....	㉓
그림 28. COIDP 도구의 형성적 측정모형(COHIP 구성개념 포함) .....	㉔
그림 29. COIDP 도구의 반영적 측정모형(COHIP 구성개념 포함) .....	㉕

## 약어 정리

- AVE: 평균추출분산(average variance extracted)
- CB-SEM: 공분산기반 구조방정식모형(Covariance-Based Structural Equation Model)
- CFA: 공통요인분석법(Common Factor Analysis)
- COHIP: 어린이 구강건강영향 프로파일(Oral Health Impacts Profile for Children)
- COIDP: 어린이 일상활동 구강영향(Oral Impacts on Daily Performances for Children)
- CR: 복합신뢰도(Composite Reliability)
- EQ-5D: EuroQol-5D
- EQ-VAS: EuroQol Visual Analogue Scale
- GoF: 적합도(Goodness of Fit)
- GSCA: Generalized Structured Component Analysis
- HRQoL: 건강관련 삶의 질(Health-Related Quality of Life)
- MSE: 평균제곱오차(Mean Square Error)
- NHIS: 국가건강면접조사(National Health Interview Survey)
- PCA: 주성분분석법(Principal Factor Analysis)
- PLS-SEM: Partial Least Square 구조방정식모형(Partial Least Square Structural Equation Model)
- QALY: 질보정수명(Quality Adjusted Life-Years)
- SE: 표준오차(Standard Error)
- SF-36: Medical Outcome Study Short Form-36
- SWB: 주관적인 안녕(Subjective Well-Being)

TTO: 시간교환법(Time Trade-Off)

VIF: 분산팽창요인(Variation Inflation Factor)

# I. 서론

## 1. 연구의 필요성 및 목적

지난 40여 년 전부터 “안녕(well-being)” 또는 ‘행복’은 삶의 질을 평가하는 중요한 화두로 다루어지고 있다. 여러 학문분야에서 삶의 질에 관한 다학제적인 연구가 이루어지고 정부에서도 주관적인 안녕(subjective well-being) 또는 행복감 정도 등을 조사하여 정책 방향에 참고하고 있으며 나아가 행복지수 등을 활용하여 국제적으로 삶의 질의 수준을 매년 비교하고 있다(윤강재, 2010).

삶의 질에 대한 정의와 이론은 매우 다양하나 가장 빈번하게 언급되고 있는 것은 Diener의 주관적 안녕 모형(Diener, 1984)과 더불어 세계보건기구의 건강에 대한 정의, 즉 “완전한 신체적, 정신적, 사회적 안녕”(World Health Organization, 1948)이다. 삶의 질을 측정하는 도구들은 대체적으로 다차원적인 구조(multidimensional construct)를 가지는 것으로 이해되고 있으며, 전반적인 안녕, 정신적 안녕, 사회적 안녕, 신체적 안녕, 영적인(spiritual) 안녕, 활동과 기능 및 개인적 환경 등의 차원들을 포함하고 있다(Linton et al, 2015).

건강에 대한 정의가 삶의 질 모형의 중요한 이론적 기반을 제공하므로 삶의 질 개념의 근간을 차지하는 것이 바로 건강관련 삶의 질(Health-Related Quality of Life, HRQoL)이다. HRQoL은 신체적, 정신적, 사회적 안녕 및 활동과 기능 등의 차원을 포함하는 다차원적인 개념으로 인식되고 있는데, 대체로 HRQoL은 일반적으로 주관적 안녕의 여러 차원 중에서 영적인 안녕과 개인적 환경의 차원을 제외한 나머지 안녕의 개념으로 정의되고 있다(Linton et al, 2015). HRQoL 측정도구는 생물학적 및 생리적 관점에서 사람의 건강을 평가하던 보건 분야의 전통적 흐름과 달리 사회학, 심리학, 경제학과 같은 사회과학의

방법론으로 접근하여 주관적인 건강상태 인식을 조사하는 방향으로 개발되었다. HRQoL은 임상적 객관적 소견이나 증상이 아니라 개인의 역할 및 기능 발휘 정도와 전반적인 안녕에 대한 본인의 느낌이나 주관적인 가치평가를 반영한다. HRQoL 도구는 지역사회 인구 대상 조사, 임상 관련 연구, 치료 효과 평가 및 보건정책의 효과 평가 등 다양한 영역에서 활용되고 있다(Kaplan et al, 1996).

HRQoL 도구는 건강의 개념을 반영하여 다차원적인 구조로 개발되었다. 개발 원리와 활용 방식 등에 따라 크게 HRQoL 프로파일 및 HRQoL 인덱스의 두 가지 종류로 구분된다. HRQoL 프로파일은 심리학적 측정원리에 기반 하여 개발되었으므로 심리측정적(psychometric) 측정도구로도 불린다. HRQoL 프로파일 측정도구는 HRQoL을 통합적으로 나타내는 단일한 통합지표 보다는 개별 차원들의 수준을 도출하고 해석하는 데 중점을 둔다. 대표적인 HRQoL 프로파일 도구로는 Short Form-36(SF-36)을 들 수 있다.

HRQoL 프로파일도구에서 HRQoL의 차원은 심리학 분야에서 널리 쓰이고 있는 구성개념(construct)<sup>1)</sup>에 준하는 것으로 간주하여 측정한다. 대부분의 심리학적 구성개념들은 구성개념이 선행하여 원인으로 존재하고 그 결과로 개별 문항, 즉 지표변수<sup>2)</sup>로 측정할 수 있는 행동이 나타난다는 기본 관계에 근거한 이론적 모형을 채택하고 이를 통계적 모형으로 구체화하고 있다. 이와 같이 구성개념이 원인이 되어 반영된 결과로 실제로 측정되는 문항의 수준이 결정되는 측정모

---

1) 구성개념(construct)이란 용어는 심리학에서 의식, 욕구, 자각, 자아 등 인간 내면의 심리적 개념들을 의미하는 것으로 시작되었으며 넓게는 다양한 제반 추상적인 개념들을 포괄한다. 구성개념은 길이나 무게 등과 같이 직접 측정하기가 불가능하므로 이를 측정하기 위한 측정이론을 확립하여야만 비로소 간접적이거나 측정 가능하다. 구성개념의 측정은 일반적으로 조작적 정의를 기반으로 선정된 관찰 가능한 다수의 지표변수(indicator)에 의해 측정하게 되는데 개별 지표변수들은 구성개념과 연관된 측면과 아울러 지표변수의 고유한 측면을 동시에 가지므로 이로 인한 측정오차가 필연적으로 발생하여 대개 구성개념의 완벽한 측정은 현실적으로 가능하지 않다. 이러한 구성개념은 오랫동안 공동체 내에서 사용되면서 구성된 것이라는 의미에서 성립되게 되었는데 심리학에서는 구성요인, 구인, 이론변수, 이론적 개념, 잠재변수 등을 동일한 의미로 사용하기도 하며 모두가 영어로 “construct”로 표현된다(이순목, 2000).

2) 지표변수는 구성개념의 원인이거나 결과를 이루는 행동이나 현상으로 구체적으로 측정 가능한 항목이다. 구성개념이 원인이 되어 그 결과로 나타나는 지표변수를 반영적 지표변수, 지표변수가 원인이 되어 구성개념을 형성하는 경우 형성적 지표변수라 부른다.

형을 반영적 측정모형<sup>3)</sup>이라 하며 반영적 측정모형에 사용되는 다수의 문항을 반영적 지표변수라고 부른다.

HRQoL 프로파일에서 다수의 차원은 HRQoL의 독특한 측면들을 반영적 측정모형으로 측정된 프로파일의 형태로 보여주며 최종적으로 각 차원 점수를 구하고 그 의미를 해석하게 된다. 일반적으로 HRQoL 프로파일의 다수의 차원 점수를 하나의 HRQoL 점수로 통합하는 방법은 공식적으로 언급되지 않는다(Ware et al, 1992).

다른 종류인 HRQoL 인덱스 측정도구는 경제학 분야의 비용-효과 분석과 의사결정이 가능하도록 건강수준을 계량화하기 위하여 개발된 HRQoL 도구들이 다(Kaplan et al, 1998). HRQoL 프로파일 도구들이 다수의 차원의 측정과 해석에 주력하는 반면, HRQoL 인덱스 도구들은 하나의 요약된 HRQoL 인덱스로 도출하는 것을 최종 목적으로 한다. 이 범주에 속하는 대표적인 HRQoL 측정도구로는 EuroQoL-5D(EQ-5D)와 Quality of Well-Being(QWB)을 들 수 있다.

HRQoL 인덱스 도구들이 단일한 지표를 생성하는 방법은 크게 통계적 방법과 선호가중치 방법의 두 가지로 분류할 수 있다. 통계적 방법은 HRQoL을 이루는 다수의 차원을 각각 단일 문항으로 측정하고 이를 적절한 가중치를 부여한 선형결합으로 산출하여 하나의 요약된 HRQoL 인덱스로 점수화하는 것이다. 여기서 다수의 반영적 지표변수로 각 차원을 측정하고자 하는 HRQoL 프로파일 도구와 달리 하나의 문항으로 하나의 구별되는 특성 또는 차원을 측정하고자 하므로 HRQoL 프로파일 도구에 비하여 일반적으로 문항의 수가 적다.

개별 문항 점수가 가중된 선형결합을 이루어 산출되는 HRQoL 인덱스는 반영적 측정모형과는 달리 개별 지표변수가 변동의 원인을 제공하고 이에 의하여

---

3) 반영적 측정모형(reflective measurement model)은 심리학 분야에서 주로 사용되는 측정모형이다. 성격이나 지능 등과 같이 개인에게 고유하게 존재하는 안정된 특성들을 측정하고자 할 때 내재된 특성이 근본 원인이 되어 그 결과 외적으로 나타나는 행동 중 대표적인 것을 지표변수로 선정하여 조사한다. 내재된 특성은 원인이 되고 조사된 지표변수를 결과로 간주하여 만드는 측정모형을 반영적 측정모형이라 한다.

구성개념이 형성되는 관계를 가지는 형성적 측정모형<sup>4)</sup>에 근거한다.

선호도에 따라 가중치를 부여하는 선호가중치(preference weight)를 이용하는 HRQoL 인덱스 산출 방법은 통계적 방법 보다 더 빈번하게 사용되었다. 선호가중치 방법은 다양한 건강상태를 건강 증상 또는 기능 이상 등 구체적인 건강조건으로 분류한 후 완전한 건강상태를 '1'점으로 정하고 죽음을 '0'점으로 하여 그 범위 내에서 현재 주어진 건강조건에 대한 조사대상자들의 선호도를 직접 조사하여 가중치로 산출하는 방법이다. 국가 간 건강수준을 비교하는 데 널리 쓰이고 있는 질보정수명(Quality Adjusted Life-Years, QALY)의 산출 시에도 이와 유사한 방법이 적용된다. 이와 같이 HRQoL 인덱스의 독특한 장점은 이상적인 건강 상태를 기준으로 현재 보고된 건강상태의 수준을 평가하여 현재의 건강 조건의 질적 수준을 계량화할 수 있다는 것이다.

HRQoL은 구성개념에 해당하는 차원들 간의 관계로 이루어져 있으므로 이러한 이론적 개념들은 구조방정식을 이용하여 경험적으로 구현될 수 있다. 앞에서 나타난 바와 같이 HRQoL 프로파일 측정도구는 심리측정의 원리에 의하여 반영적 측정모형의 원리로서 각 차원을 도출한 바 있다. 그러나 서로 다른 특성을 가지는 차원이 하나로 상위의 2차 구성개념으로 통합되는 방법은 1차 구성개념인 차원의 측정의 경우와는 상이하다. 1차 구성개념인 차원이 변화의 원인이 되어 결과적으로 반영적 지표변수의 변화를 초래하는 반영적 관계로 이루어진 것과는 달리 상위의 이차 구성개념은 하위 구성개념인 차원의 변화가 원인이 되어 그 결과로 얻어지는 '형성적인' 특성을 보이는 관계로 볼 수 있다. 따라서 HRQoL 개념의 실증적 구현을 위하여 반영적인 특성과 형성적인 특성을 모두 잘 구현할 수 있는 구조방정식 모형이 필요한 것으로 보인다.

심리학 분야에서는 전통적으로 반영적 측정모형에 의거한 구성개념이 주요

---

4) 형성적 측정모형(formative measurement model)이란 다수의 지표변수가 변동의 원인을 제공하고 구성개념이 그 결과로 형성되는 관계로 만들어지는 측정모형을 말한다. 이때 사용되는 지표변수를 형성적 지표변수라고 한다.

대상이므로 반영적 측정모형의 원리를 잘 구현할 수 있는 공분산기반 구조방정식 모형(Covariance-Based Structural Equation Model, CB-SEM)<sup>5)</sup>이 주로 사용되고 있다. 그러나 사전 연구에 의하면 CB-SEM은 형성적 측정모형의 추정에는 적합하지 않다고 한다(Hair et al, 2012). 반면 경제학과 경영학 분야에서 많이 활용되고 있는 Partial Least Square 구조방정식모형(Partial Least Square Structural Equation Model, PLS-SEM)<sup>6)</sup>의 경우 반영적 및 형성적 측정모형을 포함하는 구조방정식 모형이 용이하게 수용될 수 있다(Hair et al, 2012a). 다만 PLS-SEM은 형성적 측정모형의 특성을 기본적으로 구현하는 통계모형이므로 반영적 측정모형도 잘 구현하는 지에 관한 검증이 필요하다.

HRQoL 프로파일에 의거하여 건강수준을 평가하는 경우 기대할 수 있는 이 점은 개별 차원의 세부적인 정보를 얻을 수 있다는 장점이 있고 특정 목적의 개입이 이루어진 경우 그 예상되는 결과와 직접 관련된 차원 점수의 변화에 의하여 세밀한 평가가 가능하다는 것이다. 그러나 많은 경우 차원 단위를 넘어서 단일한 HRQoL 수준으로 요약하는 것이 반드시 필요하다. 서로 다른 집단의 건강 수준의 비교, 부작용을 동반하는 치료의 효과 평가, 보건정책의 효과 평가 등 다양한 보건 관련 국면에서 결정을 내리기 위해서는 단일한 HRQoL 요약점수가 필요하다.

만약 어떤 치료법에 대한 도입 여부에 대한 판단이 필요할 때 다수의 차원 중 일부에서는 편익이 증가하는데 다른 차원에서는 편익이 감소하는 상황이 혼재한다면 개별 차원의 증감 전체를 아우르는 HRQoL의 요약점수가 있어야 최종적으로 이 치료가 이로운 지 또는 해로운 지를 판단할 수 있다. 이러한 현실적 요

5) 공분산기반 구조방정식모형은 직접 관찰이 불가능한 추상적인 개념(구성개념 또는 잠재변수)을 모형에 포함시킬 수 있는 방법으로 확인적 요인분석과 경로분석이 합성된 통계기법이다. 반영적 측정모형의 원리를 구현하는 공분산기반 구조방정식의 구성은 지표변수를 이용하여 구성개념을 측정하는 측정모형과 구성개념 간의 관계를 규정하는 구조모형으로 이루어져 있다.

6) Partial Least Square 구조방정식모형은 직접 관찰이 불가능한 추상적인 개념을 모형에 포함시킬 수 있는 방법으로 주성분분석과 경로분석이 합성된 통계기법이다. Partial Least Square 구조방정식은 형성적 측정모형의 원리에 기반 하여 만들어졌으며 지표변수를 이용하여 구성개념을 측정하는 측정모형과 구성개념 간의 관계를 규정하는 구조모형으로 이루어져 있다.

구에 따라 HRQoL 프로파일 측정도구의 사용자들은 평균 점수나 합산 점수 등을 산출하여 HRQoL 요약점으로 사용하고 있는데, 본 연구에서 적용하고자 하는 구조방정식 모형에 따른 요약점수 산출 결과와 이러한 관행적 방법들을 비교하는 것 역시 필요하다고 사료된다.

## II. 이론적 배경

### 1. 건강관련 삶의 질(HRQoL)

#### 1.1 건강관련 삶의 질(HRQoL)

‘주관적인 안녕(Subjective Well-Being, SWB)’은 통상 ‘삶의 질’로 불리며 학계, 보건 전문가들, 및 정책 입안자들 사이에서 큰 관심의 대상이 되고 있다. 삶의 질에 대한 정의와 이론은 매우 다양하며 이는 다양한 측정도구의 개발로 이어졌다. 삶의 질에 대한 일관된 정의나 측정도구 선택 기준은 아직 부재한 상태이다. Linton 등(2015)은 체계적인 고찰을 통하여 현재 사용되고 있는 99개의 SWB 측정도구를 탐색하고 분석하였다. SWB이론에 가장 빈번하게 언급되고 있는 이론은 Diener의 SWB모형(Diener, 1984)과 세계보건기구의 건강에 대한 정의, 즉 “완전한 신체적, 정신적, 사회적 안녕”(World Health Organization, 1948)이었다. 삶의 질을 측정하는 도구들은 대체적으로 다차원적인 구조(multidimensional construct)를 가지는 것으로 이해되고 있으며, 전반적인 안녕, 정신적 안녕, 사회적 안녕, 신체적 안녕, 영적인(spiritual) 안녕, 활동과 기능 및 개인적 환경 등의 측면들을 포함하고 있다.

세계보건기구는 건강의 정의를 단지 질병이 없는 상태가 아니라 신체적, 정신적 및 사회적으로 ‘안녕’한 상태라고 정의한 바 있으므로 삶의 질과 HRQoL의 개념이 유사한 것으로 인식되었다. Schipper(1990)는 HRQoL의 개념에 대하여 신체적, 정신적, 사회적 안녕의 측면에서 개인의 경험과 신념, 기대, 인식에 따라 평가된 건강이라는 정의를 내리기도 하였다. HRQoL은 다수의 차원으로 구성된 다차원성(multidimensionality)을 보이며, 객관적이라기보다는 주관적인 것이며, 시간에 따라 변동한다는 특징이 있다(Schipper, 1996).

건강과 관련된 삶의 질을 측정하는데 초점을 맞춘 Short Form 36(SF-36), EuroQoL-5D(EQ-5D), Quality of well-being(QWB) 등의 HRQoL 측정도구들은 주로 신체적 안녕, 정신적 안녕, 사회적 안녕 및 활동과 기능의 측면을 포함하고 있다. 보건 분야에서 건강과 관련된 삶의 질 개념이나 측정도구를 활용한 연구는 1980년대 이래로 급속히 증가하여 PubMed 검색 결과 2005년에는 5000건을 상회하는 연구가, 2016년에는 30,000건에 육박하는 다수의 연구가 출판되었다(그림 1).

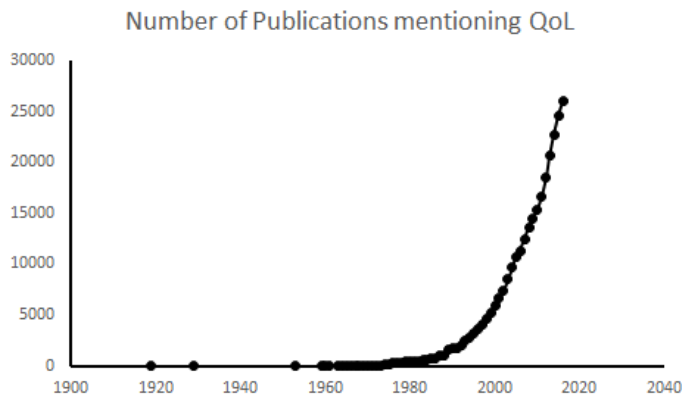


그림 1. PubMed 검색 결과 나타난 ‘건강관련 삶의 질’ 관련 문헌의 추이

## 1.2 임상연구의 결과변수로서의 HRQoL

HRQoL은 임상연구, 치료효과 연구 및 치료의 질 관련 연구에서 중요한 결과물로 사용되고 있다. 고전적인 생의학적 모형은 질병의 원인적인 인자와 병리적 과정을 중심으로 구성되었고 생물학적, 생리적 및 임상적 결과에 초점을 맞추었다. 이와 달리 사회과학의 틀을 적용한 ‘삶의 질’모형에서는 기능과 전반적인 안녕의 차원에 초점을 맞추고 인간의 복잡한 행동이나 느낌을 포괄하여 주관적인

건강상태인식에 의거한 HRQoL을 정확하게 측정할 수 있는 방법을 추구하고 있다. 건강에 대한 삶의 질 모형은 그 근원을 사회학, 심리학 및 경제학에 두고 있으므로 종종 의사나 임상연구자들에게 익숙하지 않은 개념과 방법론이 사용되고 있다(Wilson and Cleary, 1995).

Wilson and Cleary(1995)는 생의학적 모형에 나타나는 임상적 변수와 삶의 질 모형을 통합하여 하나의 모형으로 제시하였다(그림 2). 그림 2에 나타난 바와 같이 통합모형은 생물학적, 사회적 및 심리적 복잡성으로 확대해 가는 요인들을 하나의 연속선상에서 고려하고 있다.

가장 기본 단계에 위치하는 생물학적 및 생리적인 요인(Biological and physiological variables)은 가장 기본적인 건강상태의 결정인자로서 간주되며 개별적 세포, 기관 및 기관체계의 기능에 초점을 맞춘다(그림 2). 두 번째로 증상상태(symptom status)의 단계에서는 개체 전체의 반응에 초점을 맞추며 신체적 증상, 심리적 증상 등을 포함하여 비정상적인 신체적 감정적 또는 인지적 상태에 대한 환자의 인식을 관찰한다. 이 단계에서 환자가 보고하는 증상은 환자가 의료 체계에 진입하도록 하는 동기를 제공하며 이후의 의료 이용과 비용의 지출에 영향을 줄 수도 있다. 보고된 증상(reported symptom)의 결정인자로서 심리적 요인, 환자의 기대, 사회적 요인 등이 제시되고 있다(그림 2).

세 번째는 기능(function)의 단계로서 개인이 특정한 일을 수행하는 기능을 평가한다. 증상 역시 기능의 주요한 결정요인 중 하나이다. 신체적 기능, 사회적 기능, 역할 기능 및 심리적 기능 등 네 가지 차원의 기능이 포함된다. 네 번째는 전반적인 건강 인식(general health perception)으로서 이는 앞에서 나타난 모든 건강개념 및 정신적 건강을 통합하는데 주관적으로 그 등급이 매겨지게 된다. 마지막으로 전반적인 삶의 질(overall quality of life)을 삶에 대한 만족도나 행복의 정도로서 평가하게 된다(그림 2).

특기할 점은 감정적 또는 심리적 요인이 모형에 명시적으로 나타나지는 않

으나 그와 무관하게 심리적 요인은 모형의 모든 단계의 변수들과 인과적 관련성을 맺게 된다는 것이다. 예를 들면 우울감이 있는 환자는 혈당관리를 위한 노력을 하지 않게 되어 혈당치를 높이는 원인이 될 수 있고, 또한 사회적 활동을 기피하게 되어 신체적 및 사회적 기능이 나빠지는 원인을 제공할 수 있으며 또한 전반적인 건강인식을 평가하는 데 큰 영향을 줄 수 있다(Wilson and Cleary, 1995).

임상연구에서 HRQoL을 측정 결과는 대략 세 가지 측면으로 활용된다. 특정 질병 또는 건강상태와 관련된 조건의 영향을 수량화하기 위해, 임상적 개입의 결과 발생한 변화를 평가하기 위해, 임상적 치료나 건강관련 정책의 비용-효과 분석의 중심요소 등이다.

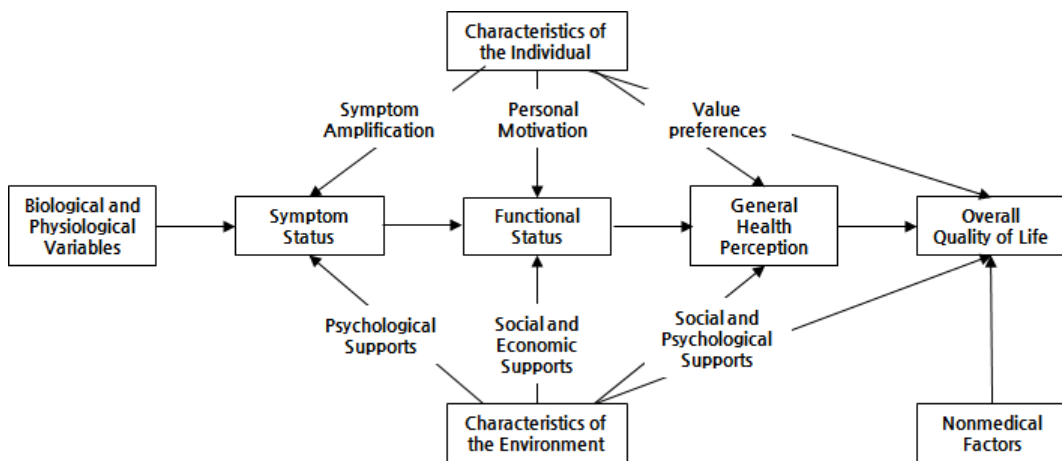


그림 2. 건강관련 삶의 질 모형에서의 환자 결과물 측정 간의 관계 (Wilson and Cleary, 1995).

### 1.3 건강관련 삶의 질 측정도구에 대한 두 접근법

HRQoL의 개념이 다차원적이라는 점에 대해서는 대체로 이견이 없으나 구체적인 측정에 있어서는 심리측정/프로파일 접근법 및 결정이론/인덱스 접근법 등 두 가지 상반된 접근법이 있고 그 결과로 HRQoL 프로파일 및 HRQoL 인덱스 측정도구가 개발되었다. 심리측정(psychometric) 또는 프로파일(profile) 접근법에서는 HRQoL의 다차원적인 특성이 측정에 있어서도 직접적으로 적용이 되어 반드시 다차원적으로 측정이 되어야 한다고 보고 이를 삶의 질의 “프로파일”로 지칭한다. 이 방법은 심리학 연구방법론의 전통을 충실하게 따르고 있으므로 심리 측정 접근법으로도 부른다.

심리측정 접근법에 따르면 오로지 그 차원에 대해서만 측정이 가능하며 차원 간의 측정은 가능하지 않다(Hunt, 1980). 이 접근법은 심리측정의 원리에 철저히 입각하여 측정도구를 개발하였으며 요인분석<sup>7)</sup>의 원리에 의거하여 각 차원을 구분하여 건강 영향의 다차원적인 프로파일을 평가하였다. 따라서 관심의 중심은 HRQoL의 전반적인 요약 보다는 HRQoL을 구성하는 다수의 차원을 평가하는 데 있다. 다차원적인 프로파일의 세부적인 정보는 특수한 상황에서 임상 연구의 내용이 구체적인 여러 측면에서의 정보를 필요로 하는 경우에 매우 유용할 수 있다.

몇 연구에서는 각 차원의 점수에 대하여 주성분분석을 하여 가중치를 결정하고 이를 크게 두 개의 상위 성분이 되는 신체적 성분(physical component score)과 정신적 성분(mental component score)의 두 점수로 요약하는 방법을 제안하기도 하였으나 다차원적인 구조의 측정도구에 의한 측정 결과를 하나의 전반적인 하나의 요약 점수로 나타내는 것은 일반적으로 제안되고 있지 않다

---

7) 요인분석(factor analysis)이란 사회과학의 측정이론 중 수량산출론에 해당되는 방법이다. 경험체계 내에서 수집된 자료는 불완전하여 지표변수의 측정에 오차가 존재하므로 측정오차를 요인분석 과정을 통하여 제거하고 순수한 구성개념에 해당하는 부분만을 추출하여 요인으로 구성하고 이를 잠재변수로 취급하는 방법이다.

(Jenkinson et al, 1997; McHorney et al, 1993; Gandeck et al, 2004).

실제로 대부분의 임상적 치료법은 치료 효과와 더불어 부수적으로 부작용을 동반하는 경우가 많다. 임상적 판단은 이러한 상반된 이점과 부작용의 영향을 종합하여 이루어져야 한다. 만약 당뇨병자에게 주사하는 인슐린 약물로 인하여 주사 부위의 피부에 문제가 발생한다고 하자. 그러면 그 문제로 인하여 그 치료를 중단하여야 하는 것인가에 대한 판단을 내릴 필요가 있다. 피부의 문제가 번거로운 것은 사실이지만 그 치료를 중단한다면 환자가 죽게 될 수도 있다. 그러므로 치료 효과와 그 부작용은 전체적인 건강의 관점에서 바라보아야 하며 궁극적으로 우리는 그 치료가 전체적으로 총 편익(net benefit)을 주는 지 전체적으로 총 손실(net deficit)을 주는 것인지 판단하여야 한다.

따라서 심리측정 접근법에 의거한 도구를 사용하는 경우에도 임상가들은 궁극적으로는 이러한 프로파일을 통합하여 전체적인 해석을 하고 임상적 판단을 하여야 한다. 예를 들어 어떤 약이 신체적 기능이나 정서적 기능 등 다수의 차원에 대하여 이러한 또는 저러한 변화를 일으켰다고 할 때 각 차원의 상대적 중요성을 판단하여 그 약의 전반적 영향을 임상적으로 판단하는 것이 현실적으로 요망된다. 그러나 공식적으로 프로파일 접근법에서는 각 차원의 평가는 가능하나 이들을 통합하는 방법은 제시되지 않는다(Ware & Sherbourne, 1992; Kaplan et al, 1998). 현실적으로 많은 임상가들은 관례적으로 상대적 중요성을 동일하다고 가정하고 가중치를 1로 부여하여 각 차원의 값의 합산점수를 요약 점수로 사용하고 있다. 그러나 이에 대한 이론적 근거의 부재는 여전히 문제로 남아 있다.

HRQoL 인덱스 도구는 임상 과정이나 특정 보건정책의 효과-비용 분석 등 판단과정에 대한 적용 가능성을 중요하게 보고 HRQoL을 여러 프로파일 이 아니라 단일한 HRQoL 인덱스로 통합하여 측정한다. 전반적인 건강의 측정 결과를 단일한 수치로 부여할 수 있다고 보아 건강의 다양한 속성에 대하여 그 영향의 중요도에 따라 가중치를 부여하고 적절한 방법론에 의하여 특정한 건강상태에 대

하여 통합적인 단일한 지표인 인덱스를 계산한다(The EuroQol group, 1990).

심리측정 접근법에 의한 삶의 질 도구가 각 프로파일에 대한 세부적인 정보를 제공한다면 HRQoL 인덱스에 의한 도구는 전체 건강에 대한 영향을 평가하여 net benefit을 제공할 수 있다. net benefit을 계산하는 데에는 복잡한 고려가 필요하다. 각 차원의 값을 합치고자 한다면 가중치 결정의 근거가 필요하다. 전통적으로 결정 이론 접근법에서는 다양한 건강의 차원에서의 결과물(outcome)에 대하여 선호 및 유용성 측정치(preference and utility measure)로 평가하여 이를 근거로 선호가중치(preference weight)를 결정하여 통합하고 있다. 선호 및 유용성 측정치는 인구집단을 대표하는 표본 집단에서 다양한 건강관련 결과물에 대한 선호 및 유용성의 정도를 평가하게 하여 그 결과를 일반화하여 사용하게 된다.

보건 프로그램의 비교에 있어서 가장 중요한 두 결과물은 생존기간과 삶의 질이다. 프로그램의 평가와 임상적 결정과정을 수행하기 위하여 삶의 질과 생존기간 간의 교환(trade-off)을 고려하여 건강결과를 하나의 점수로 도출하는 것이 필요하다. 예를 들면 보행이 어려운 건강 상태에서 10년 동안 사는 것과 완벽하게 건강한 상태로 몇 년을 사는 것이 교환 가능하다고 생각할 수 있을까? 그 선택은 개인이 보행이라는 행위에 대하여 부여하는 가치에 따라 짧거나 상당한 정도로 길 수도 있을 것이다. 이와 같이 응답자가 삶의 질과 생존기간 간의 교환(trade-off)을 고려하여 정량적으로 평가한다면 결과적으로 보행이 불가능한 삶의 가치를 평가할 수 있다. 이와 같이 개인의 선호도 또는 유용성에 대한 가치판단을 통하여 일정한 건강상태의 가치를 평가할 수 있다. 이러한 방법으로 건강상태에 대한 선호가중치를 산출하는 방법을 시간교환법(Time Trade-Off, TTO)이라고 한다(그림 3).

TTO는 죽음보다 나은 가상의 건강상태에 대하여 그 상태로 x년을 사는 것과 완전한 건강상태로 t년을 사는 것이 동일한지를 물어서 TTO를 통한 해당 건강상태의 가치를  $t/x$ 으로 추정하는 방법이다. 이때 완전한 건강상태의 가치는 최

고값 1 로 정의한다. 이 TTO 점수는 어느 하나의 건강상태에 대하여 다른 건강상태에 대한 선호의 정도를 나타낸다. 이러한 TTO점수를 여러 건강상태를 시간과 연결하는 하나의 총점으로 얻을 수 있고 이 점수는 1(완벽하게 기능하는 삶의 질) 과 0(죽음) 사이의 값을 가진다. 산출된 점수를 이용하여 QALY를 계산하거나 경제성 평가에 적용할 수 있다(그림 3).

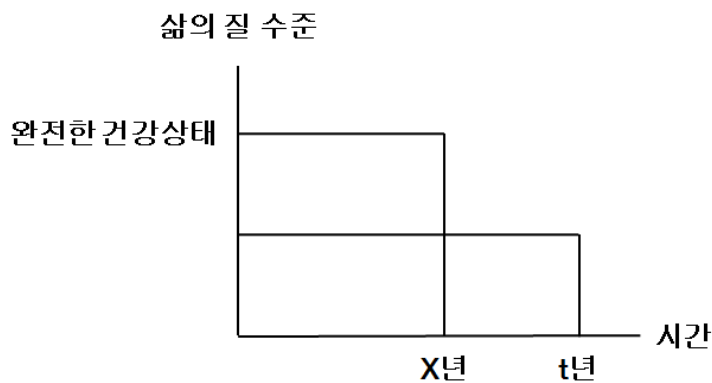


그림 3. 시간교환법(Time Trade-Off, TTO)

HRQoL 인덱스 도구들의 경우 대부분 소수의 문항으로 이루어져 있고 하나의 문항으로 하나의 차원을 측정하도록 설계되었다. 따라서 측정의 정확도와 신뢰도에는 상대적으로 부족한 면이 있다. 인덱스 도구에서 각 문항을 통합하여 하나의 HRQoL 인덱스로 통합하기 위해서 선호도기반 가중치(reference-based weight)를 사용하여 통합하고 있다. 선호도기반 가중치는 나라별 문화별 차이가 인정되므로 개별적 환경에 맞도록 각자가 별도로 개발하여 사용하고 있다.

기존 연구에 따르면 대표성 있는 표본에 대한 일회적인 조사로 선호도기반 가중치를 확정하고 유사한 인구집단에 대하여 일괄적으로 적용하는데, 이

가중치에 대한 신뢰성이나 타당성에 대한 평가방법은 존재하지 않는다. 더불어 서로 다른 조사에서 상이한 가중치가 얻어지는 경우 이를 비교하고 평가하는 기준이 마련되어 있지 않다는 문제점이 지적될 수 있다.

#### 1.4 건강관련 삶의 질 도구의 종류

현재에 이르기까지 HRQoL 프로파일과 HRQoL 인덱스에 특성을 가지는 다양한 측정도구들이 개발되어 있다(표 1)

표 1. 다양한 건강관련 삶의 질 측정도구

Type	Measures
Profile	36-item Short-Form Health Survey (SF-36)
	Sickness Impact Profile (SIP)
	Nottingham Health Profile (NHP)
	The World Health Organization Quality of Life (WHOQOL-BREF)
	Psychological General Well-Being Index (PGWB)
	Oral Health Impacts Profile for Children (COHIP)
	EuroQol (EQ-5D)
Index	Health Utility Index (HUI)
	Quality of Well-Being Scale (QWB)
	Health & Activities Limitation Index (HALI)
	Oral Impacts on Daily Performances for Children (COIDP)

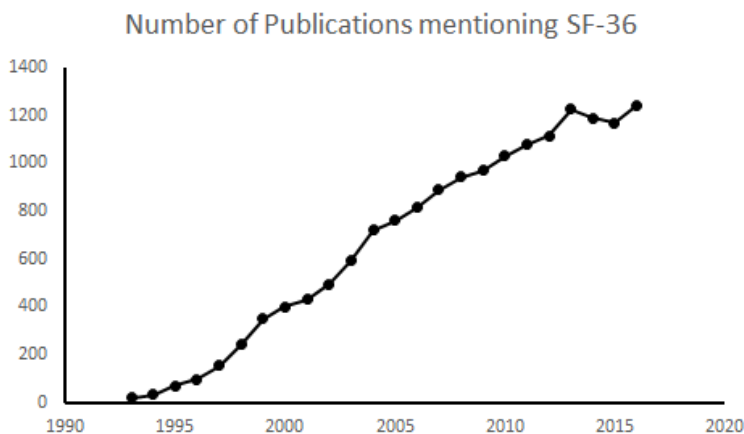
##### 1.4.1 HRQoL 프로파일 범주의 측정도구

###### 1) Medical Outcome Study Short Form-36(SF-36)

전 세계에서 가장 많이 사용되는 HRQoL 도구는 Medical Outcome Study

Short Form-36(SF-36)으로(Kaplan & Ries, 2005), SF-36 측정도구를 활용한 연구는 1990년대 이래로 급속히 증가하여 PubMed 검색 결과 2010년 이후 매년 1000건을 상회하는 연구가 출판되었다(그림 4). SF-36은 RAND cooperation<sup>8)</sup>의 Medical Outcome Study(MOS)에 의하여 개발되었다. 처음에는 RAND Health Insurance Study에서 건강 결과물 측정 전략에 기반 하여 HRQoL 측정도구의 개발이 시작되어 총 116개의 문항을 가지는 건강 관련 삶의 질 측정도구인 MOS Core Measure로 개발되었으나 이후에는 이를 20개의 문항을 가진 SF-20으로 축소하였다(Hays, 1995).

그러나 SF-20이 충분한 신뢰도를 확보하지 못하였으므로 새롭게 8개 건강 개념(신체적 기능, 신체적 역할, 통증, 일반건강인식, 활력, 사회적 기능, 정서적 역할 및 정신적 건강)을 포함하는 36 문항으로 SF-36을 재구성하게 되었다. SF-36은 훈련된 면접자가 수행하거나 응답자 본인이 직접 응답할 수도 있다. SF-36은 간단할 뿐 아니라 신뢰도와 타당도에 대한 충분한 증거를 확보하고 있다(Scott-Lennox et al, 1999).



8) RAND corporation은 1948년에 설립된 비영리 연구 기관으로서 주로 미군 관련 정책설계를 위한 연구와 개발 결정을 뒷받침하고자 만들어졌다. RAND는 “research and development”를 의미하며 현재는 연구와 분석을 통하여 공공 정책과 결정과정을 개선하는 데 주력하고 있다. RAND는 60년 이상 사실에 기반을 둔 연구와 분석을 통하여 개인, 가족 및 지역사회를 더욱 안전하고 건강하며 풍요롭게 하는 데 기여하고자 노력하고 있다(RAND Home Page, <http://www.rand.org/>).

#### 그림 4. PubMed 검색 결과 나타난 ‘SF-36’ 관련 문헌의 추이

##### 2) Nottingham Health 프로파일

Nottingham Health 프로파일(NHP)은 특히 유럽의 지역사회를 대상으로 많이 시행된 HRQoL 도구이다. NHP는 두 부분으로 구성되어 있다. 첫 부분은 수면, 신체적 활동성, 에너지, 통증, 감정적 반응, 사회적 고립 등 6개 범주에 해당하는 36개의 문항으로 구성되어 있다. 동일한 범주 내의 문항들은 상대적 중요도에 따라 평가되며 한 범주는 최저 0점 최고 100점의 점수 범위에서 측정된다.

NHP의 두 번째 부분은 건강에 의하여 가장 큰 영향을 받는 삶의 부분과 관계가 밀접한 7개의 문항으로 구성되어 있다. 그 부분들은 고용, 가족 활동, 사회생활, 가정생활, 성 생활, 취미 및 관심분야 등이다. 응답자는 건강 상태로 인하여 이런 부분에서 자신의 삶이 영향을 받는지 여부를 답변한다. NHP의 신뢰도와 타당도는 많은 연구들에서 입증된 바 있다(McEwen, 1992).

NHP는 개인과 지역사회의 건강에 대한 정의로 부터 출발한 소비자 중심의 측정도구이다. 나아가 이 척도는 지역사회 사람들이 쉽게 이해할 수 있도록 최소한의 독해력이 요구되는 언어로 만들어졌다. 오랜 연구에도 불구하고 차원 간의 상대적 중요도 가중치가 제공되지 않아서 차원 간의 비교는 대체로 가능하지 않다(McEwen, 1992).

##### 3) 어린이 구강건강영향 프로파일(Oral Health Impact Profile Children, COHIP)

COHIP는 학령기 어린이의 구강을 포함한 악안면 well-being을 평가하기 위하여 개발된 삶의 질 도구이다. COHIP는 Broder 등(2007a)에 의하여 개발되었으며 총 34개의 문항으로 구강건강, 기능적 제한, 정서적 제한, 학교활동 제한

및 자신의 이미지 등 5개 차원으로 구성되었다. 그 중 4개의 차원 28개 문항은 부정적 경험이나 인식을 질문하며 1개 차원 6개 문항은 긍정적 문항으로 구성되어 있다. COHIP 도구의 신뢰도나 타당도는 8세에서 15세에 이르는 다양한 임상 조건을 가지는 어린이들에게서 검증된 바 있다.

#### 1.4.2 HRQoL 인덱스 범주의 측정도구

##### 1) EuroQol-5D(EQ-5D)

유럽의 지역사회의 일반 인구집단에 가장 빈번하게 사용되는 HRQoL 도구는 EQ-5D로서 이 HRQoL 측정도구를 활용한 연구는 2000년대 이래로 급속히 증가하여 PubMed 검색 결과 2016년에는 800건에 육박하는 관련 연구가 검색되었다(그림 5). 이 방법은 Paul Kind 및 그의 동료들에 의해 처음 고안되어 서유럽에서 조직된 협업 집단인 EuroQol에 의해 발전되었다. EuroQol은 1987년에 영국, 핀란드, 네덜란드, 노르웨이 및 스웨덴에 위치하는 7개 센터의 구성원들로 조직되었는데 국제적인, 다수의 기관에 의한, 다학제적인 연구자들의 네트워크를 포함하였다(Kind, 1997). 최근에는 스페인을 비롯하여 독일, 그리스, 캐나다, 미국 및 일본의 학자들도 이 집단에 동참하였다. 이들의 목적은 일반적으로 통용될 수 있는 건강에 대한 평가도구를 만들어 전 유럽에서 사용 가능하도록 하는 것이었다.

EuroQol이 처음으로 개발한 측정도구는 6 개의 차원과 14개의 건강상태를 가진 형태로서 응답자들은 자신이 인식하는 건강 정도에 대하여 사망(0.0)에서 완벽한 건강(1.0) 사이의 연속적인 선상에서의 특정한 위치로 표시하였다. 이 방법은 영국, 스웨덴 및 네덜란드에서 현장 조사를 통하여 타당성을 확보하였다.

EuroQol이 발행한 가장 최근의 버전은 5개의 문항으로 구성된 EQ-5D로

다양한 임상연구 및 일반 인구집단 대상 연구에서 활발하게 이용되고 있다. EQ-5D가 간단하면서도 포괄적인 내용을 가지는 장점이 있지만 대부분의 응답자들이 최고 점수의 HRQoL 점수를 획득하는 등 천장(ceiling) 효과가 나타나는 데 대한 우려도 있다(강은정 등, 2006).

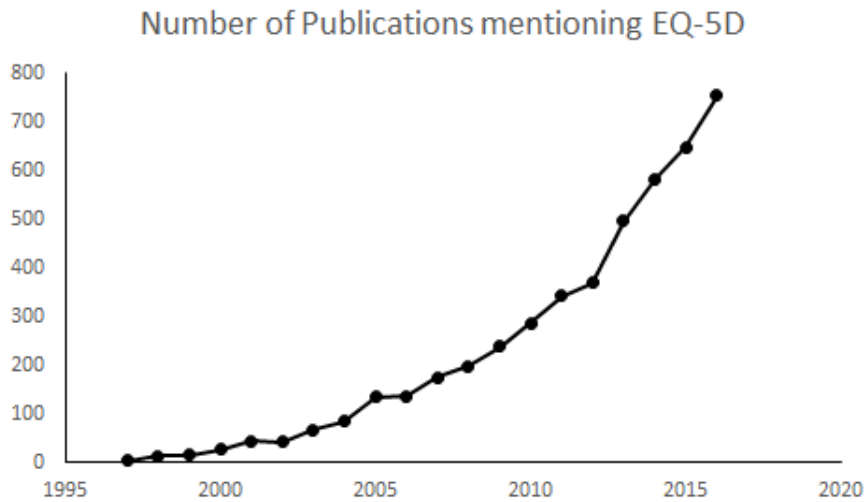


그림 5. Pubmed 검색 결과 나타난 ‘EQ-5D’ 관련 문헌의 추이

#### 4) Health Utilities Index(HUI)

HUI는 건강상태에 대한 선호도 HRQoL 도구이다. HUI는 건강 상태에 대한 분류체계를 형성 한 후, 선호도 기반 다속성(multi-attribute) 유용성 함수를 이용하여 분류된 각 속성에 가중치를 부여하고 이를 질문지에 대한 응답결과에 적용함으로써 최종적으로 HUI 점수를 도출한다. Health Utilities Index Mark 2(HUI2)는 건강상태에 대한 7개의 차원으로 구성되어 있는데 구체적 개념은 감각(시각, 청각, 말하기), 이동성, 감정, 인지, 자기 돌봄, 통증 및 생산력(fertility)이다(Feeny et al, 1995). 캐나다 일반 사회조사의 1991년 통계를 통하여 검사-

재검사 신뢰도에 관한 증거가 보고되었는데 높은 일치도를 나타내었고 이어 구성 타당도 역시 확인 및 보고되었다(Glaser et al, 1999).

#### 5) Self-Administered Quality of Well-being(QWB)

QWB 도구는 다양한 증상과 기능에 대한 선호가중치 값을 결합하는 방법으로 HRQoL을 평가한다. 선호가중치는 일반 인구집단을 대표하는 것으로 간주되는 표본 집단의 평가에 근거하여 얻어진다. 이들은 증상과 기능 등의 다양한 건강조건에 대하여 원하는 정도를 평가하여 해당 조건들을 사망(0.0)에서 완벽한 건강(1.0) 사이의 연속적인 선상에서의 특정한 위치로 표시하는 것을 가능하게 하였다. 다양한 증상 복합들이 존재하거나 부재하는 것에 대한 질문을 통하여 증상 관련 선호가중치가 얻어졌고 또한 지난 3일 간 경험한 기능적 장애를 세 개의 차원 (이동성, 신체적 활동, 사회적 활동)으로 나누어 동일한 방법으로 평가하였다. 위의 4개 차원 점수는 하나의 전체 점수로 결합되어 사망(0.0)에서 완벽한 건강(1.0)의 범위 내에서 무증상의 적정 기능에 대한 점수로 표현되었다(Kaplan & Anderson, 1996).

#### 6) Health and Activities Limitation Index(HALex)

HALex는 미국에서 국가건강면접조사(National Health Interview Survey, NHIS)와 관련하여 개발된 HRQoL 도구이다(Erickson, 1998). HALex는 NHIS에서 수집된 건강 관련 정보를 하나의 요약 인덱스로 나타내고자 하는 목적으로 개발되어 주로 Healthy People 2000 목표의 성과를 측정하는 기대 수명의 건강-가중치 요약값을 계산하는 데 주력하였다. 이 요약값은 “Years of Happy Life”으로 불리며 미국 건강통계 국가센터에서 개발하였다.

HALex는 두 개의 차원을 가진 활동성과 기능제한을 7-수준으로 분류하여 “제한 없음”에서부터 “Instrumental Activities of Daily Living(IADL)의 제한” 과 “Activities of Daily Living(ADL)의 제한” 까지 이르도록 구분하고 자가보고 전반적인 삶의 질을 excellent, very good, good, fair, poor 등 5개 수준으로 구분하였다. 또한 위의 두 구분법을 결합하여  $7 \times 5 = 35$ 개의 건강상태로 구분하였다. 이러한 건강상태의 구분에 대한 가중치는 대체로 이전에 조사된 자료에서 추정해 낸 가중치를 준용하여 적용하였다(Erickson, 1998).

이러한 방법은 다른 여러 연구에서 구한 다른 인덱스와 비교해 본 결과 일반 인구집단에서 적절한 수행을 보였다고 평가되었다. 이 측정도구의 장점은 주요 국가건강통계의 일부인 NHIS에서 계산이 가능하므로 동일하게 활동과 기능 제한 및 자가보고 건강을 조사한 다른 자료에도 적용 가능하다는 것과 충분한 량의 자료를 구하는 것이 매우 수월하다는 것이다.

#### 7) 어린이 일상활동 구강영향(Oral Impacts on Daily Performances for Children, COIDP)

COIDP 도구는 2004년 Gherunpong 등에 의하여 영어로 개발되어 스페인어, 불어, 이탈리아어, 포르투갈어, 및 탄자니아어 등으로 번역되어 타당성 검증을 거쳐 활용되고 있다. COIDP 도구는 먹기, 말하기, 치아 닦기, 잠자기, 정서적 안정, 웃기, 공부하기 및 사회적 활동 등 8개 항목으로 구성된 설문 도구로서 지난 3개월 간 경험한 구강건강의 문제로 인한 부정적 경험의 유무, 빈도 및 심도를 조사하여 구강건강 상태가 어린이의 삶의 질에 미치는 영향을 계량화 한다.

## 1.5 질보정수명을 이용한 효과 평가

하나의 통합지표로 나타내어진 HRQoL 측정 결과는 치료나 정책의 효과를 평가하는 데 활용할 수 있고 보건 분야에서 핵심적인 효과-비용 분석 연구 중 하나인 QALY를 추산하는 데 적용될 수 있다. 예를 들어 어떤 치료가 병리적 생리학적 비정상상을 치료하여 수명을 연장시킬 수 있지만 또한 삶의 질 또는 삶에 대한 만족을 감소시킬 수도 있다고 한다. 따라서 삶의 질을 고려한 상태에서 기대 수명을 보정하여 실제로 얻는 이득을 계량화하는 것이 필요하다. 이를 위하여 먼저 치료로 인하여 발생하는 다양한 건강 결과물에 대하여 선호 및 유용성 측정도구를 이용하여 HRQoL은 하나의 값으로 요약하여 그 치료로 인하여 연장 가능한 수명 기간에 대하여 감소되는 전반적 삶의 질을 고려하여 최종적으로 질보정 수명을 계산한다면 그 결과 얻는 QALY는 그 치료로 얻는 최종적인 편익이 될 것이다.

위와 같이 QALY는 임상적 개입의 효과 평가에도 이용될 수 있고 더불어 전체 인구집단의 기대 수명(life expectancy)에도 적용할 수 있다. 기대 수명에 다양한 질환의 이환 등의 건강상태를 보정한 질보정수명은 대표적인 국가 간의 건강 비교지표이다. 특정 질환과 관련한 질보정수명의 산출과정은 질병의 이환, 신체적 활동성, 사회적 활동성, 증상과 건강문제 등에 대하여 각각 가중치를 산출하고 통합하여 질병을 앓는 기간에 건강상태를 잠식하는 하나의 인덱스로 계산한 후, 이를 질병을 앓는 기간에 적용하여 건강상태를 보정한 QALY을 계산하게 된다. 결국 QALY을 계산하는 과정은 다수의 차원을 고려하는 HRQoL 프로파일에 의해서는 산출이 가능하지 않고 다차원적인 개념을 하나로 통합하는 요약점수가 필요하다.

질보정수명은 HRQoL을 이용하여 질병의 이환과 사망에 대한 정보를 결합하는 방법으로 어떤 조건의 현재 상태에 대한 영향과 더불어 미래의 수명과 기능

에 대한 영향을 통합하여 건강과 관련된 수명으로 재평가한 것이다. 예를 들면 폐암 환자의 경우 환자의 현재 기능에 미치는 영향은 감기 걸린 사람과 크게 다르지 않을 수 있으나 미래에는 폐기능에 심각한 장애가 발생하거나 조기 사망할 것을 예상할 수 있다. 따라서 건강상태의 종합적인 평가는 현재 상태 뿐 아니라 미래의 상태까지도 포괄하여야 할 필요성이 있다.

질보정수명은 보건정책의 비용-효과 분석을 하는 데 사용된다. 일반적으로 보건정책은 기대 수명을 증가시키고 사람들이 생존하는 동안 삶의 질을 증진시키는 데 그 목적이 있다. 특정한 보건정책을 실시하였을 때 기대할 수 있는 효과를 질보정수명에 의하여 평가하고 그 정책 수행에 필요한 자원을 비용으로 추산하여 비용-효과 분석을 실시할 수 있다.

더불어 이 접근법은 특정 조건에 대한 치료나 정책에 의한 개입의 이득 (benefit)과 부작용(side effect)를 함께 고려할 수 있다는 장점도 있다. 예를 들면 가려움증을 치료하기 위하여 항히스타민제를 복용한다고 할 때, 가려움증의 완화와 더불어 졸음과 무기력감이 나타날 수 있으므로 이득과 부작용이 공존하게 된다. 이 때 이득과 부작용에 각각 가중치를 주어 통합하면 net benefit을 구할 수 있고 치료 여부에 대한 판단에 참고할 수 있다.

## 1.6 정신건강과 QALY

QALY의 개념이 많은 임상 전문 분야에서 받아들여지고 있으나 정신건강 분야에서는 큰 관심의 대상이 아니다. 이는 정신건강과 신체적 건강의 결과물이 개념적으로 분명하게 구분된다는 일반적인 관념을 반영한다. HRQoL 도구가 직접적으로 정신건강 요인을 포함하고 있지 않더라도 정신건강증상은 역할기능의 질문에 반영될 수 있다. 예를 들어 우울증에 관해 본다면 역할기능이 유지되는 상태에서 우울감은 감기와 같은 심각하지 않은 것으로 볼 수 있으며 안녕에 적은

변화만을 준다고 볼 수 있다. 심각한 우울증이 있다면 그 사람은 병원이나 특수 시설에 있어야 할 것이고 역할기능의 점수가 필연적으로 낮을 것이다. 오랜 기간의 우울감의 존재는 단 기간의 우울감에 비하여 더 많은 QALY의 손실을 초래할 것이다.

비록 대부분의 HRQoL 인덱스 측정도구에는 정신건강관련 내용이 포함되어 있지 않지만 경험적 자료로 정신건강 상태와의 연관성이 유의하게 나타나 그 타당성이 입증되어 있다(Patterson et al, 1996; Rapaport et al, 1996).

## 1.7 측정 관점에서의 고찰

### 1) HRQoL 측정도구의 신뢰도

사회과학의 고전검사이론<sup>9)</sup>에 따르면 신뢰도는 검사나 측정 결과에서 전체에서 진점수가 차지하는 변동분 즉, 분산의 비율(propportion of variance) 추정치이다. 표준화한 값을 사용한 경우 1에서 이 비율을 빼면 오차 분산의 비율을 얻으며 오차분산의 값이 작아야 신뢰도가 높다. 고전검사이론에서는 잠재적인 오차의 원인으로 문항표집과 시간표집에 의한 오차에 주목하여 신뢰도를 검정한다. 문항 표집과 관련하여서는 각 지표변수가 구성개념과 관련된 무한한 동질적인 문항에서 대표적인 표본을 뽑아서 조사 문항으로 사용한다는 문항표본이론에 기초하고 있다. 따라서 문항표집과 관련하여 신뢰도란 문항 간의 상관의 정도로 추정된다.

---

9) 전통적으로 심리학 영역에서 사용되는 대부분의 표준적 검사들이 고전검사이론(Classical test theory)에 의거하여 개발되었다. 추상적인 구성개념을 측정하기 위하여 매우 큰 구체적인 행동의 집합에서 표본을 뽑아 다수의 지표변수들을 항목으로 구성하여 추상적 개념을 측정하고자 하였다. 고전검사이론에 따르면 관찰점수는 진점수와 오차로 구분되며 관련된 다수의 가정에 의거하여 이론적으로 성립한다. 고전검사이론은 반영적 측정모형을 기본으로 하여 그에 맞는 다양한 타당도와 신뢰도 기준을 마련하였다. 그러나 형성적 측정모형의 경우 고전검사이론의 기준과 맞지 않아서 적용이 가능하지 않다.

그러나 HRQoL 측정도구에 의한 건강 상태의 측정은 다른 이론적 모형을 가질 수 있다. 문항 표본들이 선정된 전체 집단이 동질적인 구성이 아니어서, 하나의 차원에서 나온 동등한 표본이라기보다는 각각이 특별한 의미를 가질 수 있기 때문이다. 예를 들면 모든 가능한 건강 증상으로부터 뽑은 표본인 두통과 요실금은 매우 다른 의미를 가지는 특성이 동등하지 않고 각자가 독자적으로 특수한 측면을 나타낸다. 이러한 경우 심리측정의 문항표본이론에 근거한 신뢰도지수 Alpha 등은 건강 상태의 측정에 적용되기 어렵다.

심리측정 이론에 의한 두 번째 형태의 신뢰도는 시간 표본이다. 심리학적 속성은 시간에 따라 변화하지 않는 것으로 간주하므로 검사-재검사를 2주 간격으로 실시한다면 측정오차를 발견하는데 큰 정보를 제공할 수 있다. 그러나 건강 상태의 경우 오늘 아프다 하더라도 2 주 후에는 완치되거나 또는 더 악화될 수도 있다. 따라서 이렇듯 기저의 구성개념이 시간에 따라 변화한다면 고전검사 이론에 의한 검사-재검사 신뢰도는 적용되기 어렵다.

## 2) 측정모형의 관점에서의 HRQoL 도구

프로파일과 인덱스 도구를 포함한 HRQoL 측정도구는 대체로 반영적 측정구조와 형성적 측정구조가 혼재되어 있는 양상으로 나타난다. HRQoL 프로파일 도구는 주로 심리측정적 방법론에 근거하여 제작 및 운용되고 있다. 대표적 프로파일 도구인 SF-36의 경우 반영적 측정모형에 의거한 차원 단위의 측정을 수행하며 차원 간의 관계에 대한 모형도 반영적인 것으로 설정하는 CB-SEM을 적용하여 분석하는 것이 일반적이다(한창완, 2009). 그러나 HRQoL의 구별되는 여러 차원들이 동질적인 성격의 전체 집합에서 표본으로 뽑힌 것으로 가정하는 문항표본이론에 근거한 반영적 관계를 가졌다고 볼 수는 없으며 각 차원들이 가지는 독특성과 상호교환 불가능성 및 차원들의 수준이 원인을 이루어 결과로서

건강상태의 수준을 결정한다는 면에서 차원들과 이차 구성개념간의 관계는 심리 측정적 이론과는 상이한 ‘형성적’ 관계일 것으로 간주할 수 있다(McHorney et al, 1993).

한편 HRQoL 인덱스 도구를 구성하는 개별 문항들은 서로 구별되는 여러 건강의 측면이나 증상 등으로 구성되어 있다. 예를 들면 EQ-5D의 경우 운동능력, 자기 관리 일상 활동, 통증, 우울 등을 각각 하나의 문항으로 측정하는데 이들은 상호 교환 가능한 특성들이 아니며 전체 건강수준은 각 문항들의 응답 수준이 통합되어 결과적으로 정해지는 등 형성적인 관계로 이루어져 있다.

이와 같이 반영적 측정구조와 형성적 측정구조가 혼재되어 있는 양상으로 볼 수 있는 HRQoL 측정도구에 대하여 각 도구의 내부구조를 구체적으로 살펴보고 그 구조적 특성을 올바르게 규정하려는 노력이 요구된다. 구조적 특성에 대한 파악이 이루어진 이후에 그 측정도구의 이론적 개념적 구조에 맞는 모형을 적용하는 것이 이루어져야 한다.

### 3) HRQoL 측정도구에 의한 점수 산출

HRQoL 프로파일 도구는 합산점수나 평균점수를 사용하여 HRQoL 점수를 산출하는 것이 일반적이다. 합산점수는 각 차원의 문항 점수에 모두 동일한 가중치 1을 부여하여 합을 구한 후 각 차원들의 점수를 다시 합하여 HRQoL 점수를 산출하는 방식이다. 이 때 차원들 간의 가중치는 대략 문항의 수와 유사하게 부여되는 것으로 볼 수 있다. 빈번하게 사용되는 평균점수의 경우 차원 점수를 문항 점수들의 평균으로 구한 후 다시 여러 차원점수들의 평균으로 HRQoL 점수를 산출하게 된다. 평균점수 사용의 경우 차원들 간을 가중치는 1이 된다.

점수를 내는 여러 대상이 반영적 관계라면 각각이 동등하다는 가정을 할 수 있으므로 가중치를 1로 부여하는 것이 합리적인 것으로 보인다. 그러나 여러 차

원과 HRQoL의 관계가 형성적이라고 본다면 여러 차원이 동일한 정도의 기여를 한다고 보기 어려우므로 적절한 가중치를 이용하여 점수를 합하는 과정이 필요하다. HRQoL 측정도구가 반영적 관계와 형성적 관계가 함께 존재한다면 이를 고려하여 도구의 특성에 맞게 규정하여 주는 것이 필요하다.

형성적 관계를 기반으로 만들어진 HRQoL 인덱스 측정도구의 경우 적절한 가중치로 통합점수를 얻는 과정을 중요시 하고 있으며, 주로 선호도 기반 가중치를 적절한 가중치로 인식하고 있다(Kaplan & Anderson, 1996). 선호도 기반 가중치는 개인의 가치평가에 의하여 얻어지므로 문화나 국가에 따라 달라 각 나라 별로 가중치를 산출하고 있다(Shafie et al, 2011; Petrou et al, 2009; Kontodimopoulos et al, 2008)) 그러나 일반적으로 하나의 대표성 있는 표본에 의하여 한 묶음의 선호도기반 가중치가 얻어지므로 선호도 가중치에 대한 신뢰성이나 타당성을 평가하고 일반화하기 어렵다는 문제점이 있다.

## 2. 사회과학에서의 측정이론의 종류

측정(measurement)이란 우리가 경험하는 사물(object)의 속성 또는 그들 간의 관계의 수준을 추정하기 위하여 일관성 있게 숫자를 부여하는 것이라고 정의할 수 있다(Stevens, 1946; Hopkins et al, 1990). 측정의 대상은 길이나 무게 등과 같이 직접 측정이 가능한 것과 인간의 잠재적 특성 등 직접 측정이 불가능하여 간접측정만이 가능한 대상으로 구분할 수 있다. 따라서 지능, 성격, 태도 등 인간의 잠재적 특성은 주로 검사라는 도구를 이용하여 간접적으로 측정하고 있다.

물리적 사물의 측정에 기반 하는 자연과학과 달리 인간행동의 측정에 기반을 두는 행동과학을 포함한 사회과학에서는 직접 측정이 불가능한 추상적 개념을 주요한 측정대상으로 설정하므로 측정 자체의 성립 여부에서부터 측정과정에 대한 논리와 방법론 등 광범위한 측면에서 이론적 체계화가 필요하게 되어 여러 갈래의 측정이론이 발전하였다. 사회과학에서의 대표적인 측정이론의 종류는 일관성 있게 숫자를 부여하는 과정을 강조하는 일관행위론, 사회과학에서의 개념의 간접적 측정과정에서 발생하는 측정오차에 주목하는 수량산출론 및 실제 측정을 가능하게 하는 기본 조건에 주목하는 현실표상론 등 크게 세 갈래로 나눌 수 있다(이순목, 2002).

### 2.1 일관행위론

일관행위론은 어떤 방법이든 관계없이 일관성 있게 사물에 숫자를 부여하는 행위 자체가 곧 측정이라고 보는 관점에서 출발하며 이는 인간 행동을 연구 대상으로 하는 사회과학에서도 자연과학에서 물리적인 자료의 측정결과를 다루는

방법을 동일하게 적용하고자 하는 측정이론의 흐름이다. 일관행위론에서는 어떤 방법으로든 일관성 있게 숫자가 뽑아졌으면 그 자체가 수량자료이므로 또 다시 수량산출을 할 필요가 없고 하나의 측정변수 자체를 하나의 이론변수로 해석한다. 일관행위론은 일찍이 1930년대에서부터 30 여 년 간 사회과학에서 주요한 측정 이론으로 활용되었는데, 기존에 도입되지 않았던 다양한 측정 대상에 대하여 일관되게 숫자를 부여함으로써 활용 가능한 정보의 폭을 넓혔다.

일관행위론의 대표적 측정방법은 사람이 어떤 사물에 대하여 인지하는 속성을 반영하여 이를 일정한 점수로 표현하는 평정척도(rating scale)로서 Likert 척도 또는 1에서 10점 등 일정한 범위를 정하여 대상자가 적절하다고 생각하는 점수를 선택하도록 하는 것이다. 일관행위론에서는 평정척도에서 산출되는 숫자를 그대로 수량척도 수준으로 간주하고 자연과학에서 적용하고 있는 다양한 통계적 분석방법을 적용하여 사회과학의 연구의 활성화에 큰 기여를 한 바 있다.

그러나 일관행위론에 의거한 측정의 관행은 현실표상론에 의하여 측정 행위 자체가 성립성에 대한 근본적인 규명이 없이 검사를 사용하는 행위라는 근본적인 비판을 받았고, 또한 일차적인 경험체계에 부여된 숫자가 어떤 속성에 대한 오차가 포함된 자료이므로 직접적으로 이론변수로 간주하는 것은 비현실적이라는 수량산출론에 의한 문제 제기에도 직면하고 있다. 이에 따라 일관행위론은 점차 방법론적으로 수량산출론에 수렴하는 경향을 보이고 있으며, 1960년대부터는 문항반응이론 발전과 더불어 검사자료에 문항반응이론을 적용하여 분석하는 경향이 우세하게 되었다(이순목, 2002).

## 2.2 수량산출론

수량산출론은 사회과학적 이론 또는 관찰을 통한 증거에 기반을 두어 사물에 수량적 속성이 이미 존재한다고 가정한다. 따라서 속성이 수량적이라는 가정은

이론적 내용의 적절함이 중요한 기본조건이 된다. 수량산출론의 적용 대상이 되는 사물의 주요 속성은 행동과학에서 인간의 행동의 원인이 되는 이론변수 또는 구성개념이며 기본적으로 연속적인 값을 가지는 수량으로 표현될 수 있다고 가정한다. 수량산출론에서 말하는 측정의 개념은 이론변수와 측정된 자료의 관계, 또한 이론변수 간의 관계 등을 설명하는 모형을 구성하여 이를 경험체계에서 관찰된 자료에 적합 시킴으로써 해당 모형에 의거하여 경험자료 속에 내재된 이론변수의 수준 또는 양을 재는 것이다.

수량산출론은 경험체계 내에서 1차로 수집된 자료는 불완전하여 자료에 측정의 오차가 있다고 보아 일차적으로 요인분석 모형과 같은 수량산출모형을 적용시켜 측정의 오차를 제거한 후, 측정하고자 하는 속성에 해당하는 부분에 대한 수량 값을 뽑아 이론변수의 측정값으로 구성한다. 이론에 맞는 통계모형을 적용하여 어떤 양이 얼마나 있는지, 즉 오차를 제외하고 난 후 사물의 속성수준이 얼마나 있는지를 알아낸 후 비로소 이론변수 간의 관계를 밝혀내는 이론모형을 구성한다.

수량산출모형은 경험 자료와 잘 맞아야 하므로 모형과 자료의 합치도(fitness)를 통계적인 확률 개념에 의거하여 계산하고 가장 큰 합치도(fitness)를 보이는 모형을 선택하게 된다. 이와 같이 수량산출론은 행동과학에 오차와 확률의 개념을 도입하여 이론변수를 측정하고 이론변수간의 관계를 밝혀내는 큰 공헌을 하였다. 수량산출론의 맥락에서 이루어진 중요한 성과는 요인분석, 다차원척도법(multidimensional scaling) 및 문항반응이론 등이 있다. 현대의 행동과학이 계량적 과학의 하나로 인식되는 근거는 대부분 수량산출론의 이론과 방법론에 의거한 것이다(Michell, 1986).

### 2.3 현실표상론(Representational Theory)

현실표상론은 실제 측정을 가능하게 하는 기본 조건으로서 경험체계 내에 수리체계로 표상(representation)할 수 있는 부분 또는 성질이 과연 존재하는 것인가의 확인을 하는 것이 선행되어야 한다는 관점에서 출발한다(Stevens, 1946). 현상표상론은 현실의 경험체계에서 얻은 수리적 자료가 어떤 조건을 만족시킬 때만 측정으로 인정할 수 있다고 가정한다.

현실표상론에서 가정하는 조건이란 구체적으로 경험체계 내의 관계가 수리체계로 표상될 수 있을지를 검토하는 데 사용하는 조건들로서, 수학적인 용어로서는 공리(axiom)라고 부른다. 이와 같이 현실표상론은 공리체계 중심(axiomatic emphasis)의 측정이론으로서 ①사물의 구분, ②사물간의 순서, ③두 사물간 차이의 비율, ④두 사물간 크기의 비율의 네 가지 성질에 대한 공리체계에 근거하고 있다.

현실표상론이 응용연구자들에게 요구하는 것은 경험체계에 숫자를 매김에 있어 먼저 경험체계 내의 어떤 특성의 성질 또는 관계를 표상할 것인지 생각해 보고 적합한 표상이 되도록 숫자를 매겨야 한다는 것이다. 현실표상론에서 시작된 Stevens의 측정수준의 개념은 자연과학과 사회과학의 수량적 연구방법론의 기본이 되었으며, 또한 현실표상론에서는 ‘연결측정구조’라는 개념을 제시하여 물리적 사물관계에서와 같이 이음작업을 필수로 하는 연장구조가 아니더라도 사회과학에서 등간척도 수준의 기본적 측정을 가능하게 하는 큰 공헌을 하였다. 그러나 현실표상론의 공리체계에 의한 엄격한 평가기준에 의해 파생하는 현실적 문제점의 하나로 응답자의 자료에 오차가 있을 것을 고려하여 확률적으로 의사결정을 하는 오차이론이 허용되지 않는다는 점이 지적되고 있다.

### 3. 자연과학과 사회과학에서의 측정오차

측정오차란 측정값과 참값의 차이이며 측정오차의 절대값을 오차의 크기라고 한다. 오차의 크기가 작을수록 측정이 정밀한 것으로 간주한다. 참값을 알고 있는 경우 참값의 크기에 대한 오차의 크기를 퍼센트로 나타낸 것을 상대오차라고 한다.

#### 3.1 자연과학에서의 측정오차

자연과학에서 물리적인 사물에 대한 측정을 할 때 발생하는 측정오차의 종류는 계통오차, 과실오차, 우연오차 등의 세 종류로 나누어진다.

##### 3.1.1 계통오차(systematic error)

계통오차는 명확한 원인에 의하여 일정 값 만큼 참값과 달라지는 등 일관성 있게 발생하므로 오차의 부호와 크기를 추정할 수 있고 따라서 이를 적절히 보정하여 바로잡을 수 있다. 계통오차는 다음 세 종류로 나눌 수 있다.

첫째, 측정계기의 불완전성으로 인하여 발생하는 계기오차이다. 예를 들면 온도계, 계기판 등의 눈금이 잘못 표기되었거나 영점조정이 안된 경우에 일관된 오차가 발생한다. 둘째, 측정 시의 온도, 습도, 압력 등 외부 환경의 영향으로 생기는 환경오차이다. 환경 오차의 예로는 측정기구의 온도에 따른 팽창과 수축으로 인한 눈금의 변화에 따라 오차가 발생하거나 질량 측정 시 공기의 부력에 의한 영향으로 질량이 과소 또는 과대하게 측정되는 경우 등이 있다. 셋째, 관찰대상의 변동에 따른 오차이다. 만약 대상의 특성이 시간에 따라 변화한다면 측정하는 시간에 따라 값이 변동하게 될 것이다. 예를 들면 일정한 주기로 또는 낮과 밤에 따

라 달라지는 측정치의 경우에는 측정 시점에 따라 측정된 값이 일정하지 않게 된다.

### 3.1.2 과실오차(erratic error)

측정자가 부주의하여 발생하는 오차를 말한다. 관찰자가 척도의 숫자를 잘못 읽거나 계산 오류 등으로 발생하는 오차를 말한다. 측정자가 충분히 주의를 기울여 제거할 수 있다.

### 3.1.3 우연오차

주위의 사정으로 측정자가 주의해도 피할 수 없는 불규칙적이고 우연적인 원인에 의하여 발생하는 오차로서 보정할 수 없는 오차이다. 원인이 불분명하며 일반적으로 개별 원인에 의한 기여는 매우 작으나 전체적으로 측정값에 불규칙하게 영향을 미친다. 우연오차 또는 랜덤오차(random error)라고 한다.

경험적으로 우연오차는 다음과 같은 성질이 있다고 여겨진다. 첫째, 우연오차는 양과 음의 오차는 같은 확률로 나타나 평균값은 0이다. 둘째, 우연오차의 크기가 작은 오차가 발생하는 빈도가 높은 오차가 발생하는 빈도보다 높다. 셋째, 크기가 매우 큰 오차가 발생할 가능성은 거의 없다. 따라서 전통적으로 우연오차는 정규분포의 형태와 유사한 것으로 여겨지고 있다(Rosner, 2006).

계통오차나 과실오차의 경우 그 계기나 환경 및 대상의 조건을 적절히 반영하여 보정할 수 있으나 우연오차의 경우 보정할 수 있는 방법이 없으므로 적절하게 그 분포를 가정하고 모형에 반영하는 방식으로 처리하는 것이 일반적이다.

### 3.2 사회과학에서의 측정오차: 특수오차(specific error)

인간행동의 측정에 기반 하는 행동과학을 포함한 사회과학에서는 직접 측정이 불가능한 추상적 개념을 주요한 측정대상으로 설정하게 된다. 우울증을 예로 들면, 우울증 자체는 개념적인 것으로 하나의 경험적 실체임은 분명하지만 인간의 정신과정에서 무형적으로 발생하므로 우울증 그 자체를 직접적으로 또는 물리적으로 측정할 수 있는 방법은 없다. 따라서 우울증이 원인이 되어 현상으로 나타나는 관찰 가능한 사람의 행동이나 본인의 느낌을 통하여 간접적인 방법으로 우울의 정도를 측정한다.

우울증이 원인이 되어 나타날 수 있는 사람의 느낌이나 행동은 매우 많을 것인데 그 전체 중 대표적으로 나타날 것으로 보이는 표본을 뽑아 그 내용을 질문하는 측정도구를 만들게 된다. 예를 들면 ‘의욕이 없고 슬프다’거나 ‘갑자기 울거나 울고 싶을 때가 있다’ 등은 우울증에 의해 나타나는 느낌을 나타낼 수 있으며, ‘잠을 잘 못 잔다’ 라거나 ‘체중이 준 것 같다’ 또는 ‘변비가 있다’ 등은 우울증의 결과 나타날 수 있는 신체적 증상을 나타낸다고 보아 이들과 유사한 대표적인 느낌이나 행동을 표본으로 뽑아서 조사한 결과에 따라 우울증 정도를 판단하는 것이다.

그러나 조사된 느낌이나 행동은 내용적으로 일정 정도 우울증에 의하여 유발된 부분이 있으나 또한 그 일부는 해당 조사항목의 독특한 특성에 의하여 발생된 부분이 같이 포함되어 최종 조사 결과로 나타날 것이라고 보는 것이다. 이러한 오차는 우연이 아니라 체계적으로 발생하게 되므로 우연오차 혹은 랜덤 오차와 구분하여 특수오차(specific error)라고 부른다.

따라서 사회과학에서의 측정오차는 자연과학에서의 측정오차의 개념과 근본적으로 다른 면이 있다. 측정하고자 하는 개념을 직접적으로 측정하지 못하고 간접적으로 측정하므로 실제로 측정된 값은 참값에 앞에서 설명한 우연오차에 더하

여 해당 조사 항목이 가지는 특수오차(specific error)까지 더해진 결과 생성된 값이 되는 것이며 이를 고유오차(unique error)라고 한다.

따라서 사회과학에서 직접적으로 측정이 불가능한 구성개념의 통계적 처리를 할 때에는 특수오차와 우연오차를 모두 고려하여야 한다. 개념적으로 사회과학에서의 측정변수의 분산은 구성개념에 대한 기여분을 의미하는 공통분산(common variance)과 다른 지표변수들과 공유하지 않고 해당 변수에만 고유한 고유분산(unique variance)으로 나누어진다. 고유분산은 변수의 특수한 측면을 반영하여 규칙적이고(systematic) 신뢰도가 높은 특수분산(specific variance)과 랜덤하며 신뢰도가 낮은 오차분산으로 구분된다.

이러한 오차에 대한 인식은 고전검사이론(Classical test theory)의 근간을 이루어 진점수와 오차가 구분되고 이를 바탕으로 신뢰도와 타당도에 관한 많은 이론과 방법론이 발전되었다. 그러나 고전검사이론의 범위에서는 체계적으로 발생하는 특수오차를 수리적으로 처리해 낼 수 있는 양적 방법론에는 이르지 못하였다.

이후 고유오차의 개념은 사회과학의 측정이론 중 수량산출론에서 요인분석(factor analysis) 또는 요인분석법이 확장된 형태인 CB-SEM의 통계적 모형으로 적용되어 경험적 모형으로 구현되었다. 요인분석 및 CB-SEM 모형에서는 지표변수 등 경험체계 내에서 구성개념을 반영하기 위하여 1차로 수집된 자료의 관찰값은 순수한 구성개념에 기여하는 부분과 고유오차를 포함하는 측정 오차로 구분된다고 간주한다. 따라서 요인분석 모형과 같은 수량산출모형을 적용시켜 측정 오차를 제거하고 순수한 구성개념에 해당하는 부분의 변량만을 추출하여 모형을 구성하는 독특한 방법론이 발전하게 되었다.

## 4. 구성개념과 측정모형의 구분: 형성적 및 반영적

### 4.1 구성개념의 정의

구성개념은 영어로는 “construct”이며 행동과학에서 사용되는 이론변수, 잠재변수, 구성개념 및 구성요인 등을 모두 포괄한다. 이 개념들은 이론적 구조에서 핵심적인 개념으로 사용되므로 이론변수라고 하며 직접 관찰하는 것이 아니라 측정변수 속에 들어 있는 알맹이로 간주되므로 잠재변수라고 한다. 이 개념은 손에 잡히는 물리적 개념에 비하여 오랜 기간 동안 공동체 내에서 사용되면서 구성된 것이므로 구성개념, 구성요인 등으로 불리게 되었다(이순목, 2000).

### 4.2 반영적 구성개념과 반영적 측정모형<sup>10)</sup>

반영적 구성개념은 잠재요인이 원인이 되어 관찰 가능한 구체적 행동, 특성 또는 그와 관계된 변수들의 수준에 영향을 미치는 구성개념을 지칭한다(Fornell 과 Bookstein, 1982). 반영적 측정모형은 반영적 구성개념을 관찰되는 반영적 지표변수의 원인으로 설정하므로 반영적 구성개념의 변동이 관련된 모든 반영적 지표변수의 변화를 일으키게 된다.

반영적 구성개념은 관찰이 불가능하더라도 그 자체가 독립적인 개념으로 연구자가 인지하고 해석할 수 있는 명백한 현상이다. 반영적 구성개념의 고전적인 예는 ‘성격’으로, 성격이라는 보이지 않는 특성에 따라 구체적인 행동이 서로 다르게 나타난다. 따라서 반영적 측정모형에서 관찰 가능한 반영적 지표변수들은 반

---

10) 측정모형이란 직접 측정이 불가능한 추상적 구성개념을 측정하기 위하여 직접 관찰이 가능한 지표변수를 이용하여 구성개념의 수준을 추정하는 모형을 만든 것이며, 반영적 측정모형은 구성개념의 수준을 반영하여 지표변수의 수준이 결정되는 관계를 모형화한 것이고, 형성적 측정모형은 지표변수의 수준이 반영된 결과로 구성개념의 수준이 ‘형성’된 관계를 모형화한 것이다.

영적 구성개념의 함수로 표현된다.

Rossiter(2002)는 반영적 구성개념과 유사한 잠재 속성(eliciting attribute)이라는 개념을 제시하면서 잠재 속성이란 사람에 따라 그 특성이 다르게 인지되며 내면적인 trait 나 state에 해당한다고 규정하면서 잠재속성과 관련된 지표변수는 명료한 활동으로 근접한 결과 행동(proximal consequences)이라고 지칭하였다.

반영적 지표변수는 동일 반영적 구성개념과 관련된 균질한 특성을 가지는 무한하게 커다란 결과행동의 집합 중에서 그 집단을 대표하기 위하여 무작위적으로 뽑힌 특정 결과행동으로 구성된 것이라고 볼 수 있으며 이는 영역 표본 이론(Domain sample theory)으로 불리고 있다. 따라서 반영적 지표변수는 하나의 반영적 구성개념에 관련된 단일차원(unidimensional)으로 이루어져 있으며 동질적인 특성을 띄므로 상호 교환이 가능하다. 또한 하나의 차원에서 추출된 다수의 반영적 지표변수가 측정하고자 하는 참값은 상호연관성이 매우 클 것이다.

실제 관찰된 값이 참값을 잘 반영한다면 반영적 지표변수의 값들은 큰 상관을 보여야 신뢰할 수 있다고 인정받을 수 있다. 이는 내적신뢰도라는 개념으로 측정되며 바람직한 신뢰도 기준으로는 대략 alpha값 0.8이 적절하다 (Nully & Bernstein, 1994).

반영적 구성개념은 종종 척도(scale)이라고도 불리며 반영적 측정모형으로 구체화된다. 반영적 측정모형은 구성개념인 요인과 측정되는 지표변수로 구성되며 그림 6으로 표현할 수 있다.

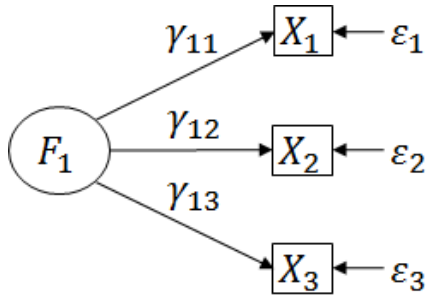


그림 6. 반영적 측정모형

$X$ 는 측정변수인 반영적 지표변수,  $F_1$ 은 반영적 잠재변수 또는 요인,  $\lambda$ 는 지표변수가 요인에 기여하는 정도인 요인계수(factor loading)를 의미하며  $\varepsilon$ 는 측정오차를 의미한다. 잠재변수와 측정오차는 서로 독립이다. 이를 행렬식으로 표현하면 식(2.1)과 같다.

$$X = \Lambda F_1 + \varepsilon \quad (\text{식 2.1})$$

벡터  $X$ 는  $n \times 1$ 의 형태로 측정변수를 나타낸다. 이는 요인계수로 불리는  $\Lambda$ 와 요인  $F_1$ 의 곱과 측정오차의  $\varepsilon$  합으로 구성된다. 벡터  $\Lambda$ 는  $n \times p$ , 벡터  $F_1$ 는  $p \times 1$ , 벡터  $\varepsilon$ 는  $n \times 1$ 의 형태를 가진다. 여기서  $n$ 은 측정변수의 수를 의미하고  $p$ 는 요인의 수를 의미한다(이순목 등, 2011).

#### 4.3 형성적 구성개념과 형성적 측정모형

형성적 구성개념이란 관찰된 구체적 행동, 특성 또는 그와 관계된 측정 수준

이 선행하는 원인이 되어 그 결과로 형성되는 구성개념으로 지표변수들에 의하여 형성되거나 유도되는 구성개념이라는 의미이다(Bagozzi & Fornell, 1982; Edwards and Bagozzi, 2000). 사회경제적 지위(socioeconomic status)가 직업, 교육 및 소득 등에 의하여 정의되는 것과 같이 구성개념이 특정한 성분 변수들의 복합으로 인식되는 경우 이를 형성적 구성개념이라고 지칭한다(Marsden, 1982).

형성적 구성개념을 측정하는 형성적 측정모형은 관찰된 행동 특성 등을 표현하는 지표변수의 가중치를 부여한 선형 결합(linear combination based on weight scores)으로 표현될 수 있으며 지표변수의 변화에 의하여 형성적 구성개념의 수준이 결정되거나 변화한다. 그 역의 경우는 성립하지 않으므로 형성적 구성개념의 수준이 변화하였다고 하여 개별 지표변수의 변화를 일으키는 원인으로 작용하지 않는다.

즉 형성적 구성개념의 수준 변화는 특정한 지표의 값이 달라짐에 따라 그 결과로 발생할 뿐이며, 형성적 구성개념의 수준이 변화하였다고 해서 해당 특정 지표 이외의 다른 지표들의 수준이 달라지지 않는다. 형성적 구성개념의 측정에 사용되는 지표변수들을 형성적 지표변수라고 부른다.

형성적 구성개념에 대한 위와 같은 관점은 비교적 최근에 Marketing 분야에서 측정도구의 개발과정을 제안한 바 있는 Rossiter(2002)가 제안한 형성 속성(formed attribute)과 유사한 것으로 보인다. Rossiter(2002)에 따르면 형성 속성은 여러 측면을 가지고 있으므로 그 중 사람들이 어떤 측면을 중점적으로 경험하였는지에 따라 그 속성을 서로 다르게 인지할 수 있으며 그러한 특성들이 합쳐져서 형성속성을 이루게 된다. 형성 속성들은 서로 다른 독특한 특성을 가지는 속성들이므로 반영적 속성의 경우와는 달리 상호 교환은 가능하지 않다. 형성속성은 여러 구성요소로 복합적인 형태로 이루어지므로 복합속성으로도 지칭할 수 있다.

Rossiter(2002)는 형성속성의 예로 서비스의 질을 거론하였는데 서비스의 질은 여러 특수한 활동의 합으로 측정 가능하다고 보았으며 이러한 활동들은 '지정

시간 내 배달' 과 같이 누가 보아도 명백하여 일관성 있게 측정 가능한 활동이어야 한다고 하였다. 선행된 행동에 의하여 결과적으로 형성속성의 수준이 정해지므로 그는 이러한 선행된 행동을 근접 선행행동(proximal antecedents)이라고 표현하였고 이 선행행동이 각각의 지표변수가 되며 형성속성은 지표변수의 가중평균합인 인덱스로 표현된다고 하였다.

형성적 구성개념에는 원인 지표변수(causal indicator) 또는 복합 지표변수(composite indicator)를 갖는 두 가지 종류의 형성적 측정모형이 적용될 수 있다. Bollen 과 Bauldry(2011)는 복합지표변수만을 특별히 형성적 지표변수로 지칭하기도 하였다. 원인 지표변수 또는 복합 지표변수를 가지는 형성적 구성개념은 다음과 같이 그림 7로 표현할 수 있다.

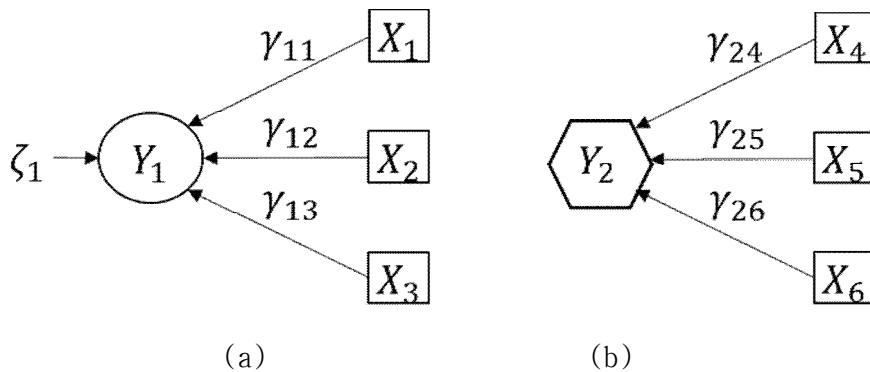


그림 7. 형성적 측정모형의 두 종류: (a) 원인 지표변수(causal indicators), (b) 복합 지표변수(composite indicators)

X는 측정변수인 형성적 지표변수,  $Y_1$ 과  $Y_2$ 는 형성적 잠재변수 또는 요인,  $\lambda$ 는 지표변수가 요인에 기여하는 정도인 인덱스 가중치(index weight)를 의미하며  $\zeta$ 는 측정오차를 의미한다. 지표변수와 측정오차는 서로 독립이다.

그림 7-a의 원인 지표변수 기반 형성적 측정모형 및 그림 7-b의 복합 지표변수 기반 형성적 측정모형을 행렬식으로 표현하면 각각 식(2.2) 및 식(2.3)과 같다.

$$Y_1 = \Lambda X + \zeta \quad (\text{식 2.2})$$

$$Y_2 = \Lambda X \quad (\text{식 2.3})$$

벡터  $Y$ 는 형성적 잠재변수로서  $p \times 1$ 의 형태로 나타난다. 이는 인덱스 가중치로 불리는  $\Lambda$ 와 지표변수  $X$ 의 곱과 측정오차의  $\zeta$  합으로 구성된다. 벡터  $\Lambda$ 는  $p \times n$ , 벡터  $X$ 는  $n \times 1$ , 벡터  $\zeta$ 는  $p \times 1$ 의 형태를 가진다. 여기서  $p$ 는 요인의 수를 의미하고  $n$ 은 측정변수의 수를 의미한다. 모든 지표변수와 측정오차 간의 무상관을 가정한다(Sarstedt et al, 2016).

원인적 지표변수 기반 형성적 측정모형에서 측정오차의 존재는 현재의 지표변수들 이외에 구성개념의 주요한 측면을 대표하는 지표변수의 누락 가능성을 고려하는 것이라고 볼 수 있다. 복합적 지표변수 기반 형성적 측정모형에서는 오차항을 배제하고 적절한 가중치를 부여한 복합 지표변수들의 선형결합인 복합지수(composite index)를 구성개념의 대리자(proxy)로 사용한다(Sarstedt et al, 2016).

원인 지표변수와 다르게 복합지표변수의 측정모형에는 오차항이 없는데 이는 주성분 기반의 형성적 측정모형의 개념규정에서 중요한 의미를 내포한다. 즉, 여기서 사용된 복합지표는 구성개념의 진정한 원인(truly causing)으로 작용한다고 보다는 구성개념에 대한 기여자(contributor)로 기능한다고 보아야 하며(Sarstedt et al, 2016), 복합지표변수의 선형결합으로서 구성개념을 대리한다.

복합 모형과 원인적 모형의 이와 같은 차이가 있음에도 과거에는 많은 연구자들이 동일시하여 사용하는 경향이 있었으나 최근에는 연구자들 사이에서 점차 구별의 당위성이 인식되고 있다(Bollen et al, 2011). 실제 모형 구현의 통계적 과정에서 구성개념이 조작적인 정의를 통하여 이론적 변수의 대리로 간주된다는 관점에서 본다면 복합 지표변수의 선형결합인 복합지수를 구성개념의 대리로서 사용하는 것은 논리적으로 타당하며 현실적 적용이 편리한 장점이 있다(Sarstedt

et al, 2016).

#### 4.4 반영적 및 형성적 지표변수의 예

반영적 및 형성적 지표변수의 예로서 호텔에 대한 만족도를 측정하기 위하여 각각의 측정모형을 적용하는 경우를 고려하여 보자. 반영적 지표변수로 측정모형을 구성한다면 다음과 같은 반영적 지표변수를 예시할 수 있다.

- 대체로 나는 이 호텔에 만족한다.
- 나는 이 호텔에서 편안함을 느낀다.
- 나는 이 호텔에 감사한다.
- 나는 이 호텔에서 묵고 싶다

호텔에 대한 만족도를 측정하기 위한 형성적 지표변수로는 다음의 질문들을 생각해 볼 수 있다.

- 방의 가구가 좋다.
- 근무자들이 친절하다.
- 방이 조용하다.
- 방이 깨끗하다
- 호텔의 서비스가 좋다
- 호텔의 가격이 저렴하다
- 호텔의 식당이 좋다

아래 표 2는 반영적 및 형성적 지표변수에 맞는 측정모형의 선택 기준과 판

단 결과를 나타낸다. 지표변수와 구성개념 사이의 인과적 관계의 방향, 지표변수의 결합으로 구성개념이 변화하는 관계인지 여부, 지표변수들의 교환가능성 등 주요 지표변수와 구성개념 사이의 내용적인 면을 이론적으로 고려하여 반영적 또는 형성적 측정모형을 선택하여야 한다. 내용적 선택기준 이외에 양적인 선택기준을 마련하고자 하는 노력이 있으나 아직 일반적으로 받아들여지는 명료한 기준은 존재하지 않는 것으로 보인다(Gudergan, 2008; Bucic, 2004).

표 2. 측정모형 선택에 대한 기준과 판단(김장현 등, 2014)

기준	판단 결과	참고
	구성개념으로부터 지표변수	
지표변수와 구성개념 사이의 인과적 관계	로: 반영적 지표변수로부터 구성개념으 로: 형성적	Diamantopoulos & Winklhofer(2001)
구성개념이 다수의 지표변수들의 결합인가 또는 설명하는 특성인가?	만약 특성이라면: 반영적 만약 결합이라면: 형성적	Fornell & Bookstein (1982)
지표변수들이 구성개념에 대한 결과나 원인을 표현하는가?	만약 결과라면: 반영적 만약 원인이라면: 형성적	Rossiter (2002)
구성개념의 특성이 변화함에 따라 모든 지표변수들이 유사한 정도로 변화하는가?	만약 맞다면: 반영적 만약 아니라면: 형성적	Chin (1998)
항목들은 서로 교환가능한가?	만약 맞다면: 반영적 만약 아니라면: 형성적	Javis, MacKenzie, & Podsakoff (2003)

#### 4.5 반영적 측정모형과 형성적 측정모형의 차이점

반영적 측정모형과 형성적 측정모형은 지표변수의 특성이 서로 다르고 구성개념과의 관계가 상이함에 따라 지표변수 간의 상관과 상호교환성, 모형 식별 등의 여러 면에서 차이점을 보이게 된다.

##### 4.5.1 지표변수 간의 상관

반영적 측정모형에서 하나의 반영적 구성개념과 연관된 여러 현상적 지표변

수들은 공통적으로 그 구성개념의 영향을 받아 변화하는 내생(endogenous) 변수이므로 지표변수들 간의 관계는 강한 양적 상관관계를 가져 상관계수가 양의 큰 값으로 나타난다.

차원 표본이론에 따르는 반영적 측정모형에서는 동일한 구성개념으로부터 영향을 받는 동질적인 결과행동의 무한 집합 중에서 무작위 표본으로 추출하여 지표변수를 구성한 것으로 간주하기 때문이다. 따라서 모집단에서의 관계를 대표하는 지표변수들 간의 상관은 상당히 커야 한다.

이는 내적 일관성의 개념으로 반영적 측정모형의 중요한 원리로 받아들여지고 있다. 만약 반영적 측정모형에서 지표변수들 간 상관이 상당히 낮다면 신뢰도가 낮은 것으로 간주하여 근본적인 검토가 필요할 수 있다. 반영적 모형에서는 각 지표변수들이 상호 수렴하는(convergent) 관계를 기대할 수 있다.

반면 형성적 측정모형에서 관찰되는 지표변수는 서로 독립적으로 구분되는 측면(facet)을 나타내는 외생(exogenous) 변수이다. 그러므로 형성적 지표변수들 간의 상관관계는 경우에 따라 매우 약하거나 또는 강할 수도 있고 양의 방향, 음의 방향 또는 무관할 경우 등 다양하게 나타날 것이다. 만약 특정 형성적 지표변수들 간에 매우 높은 상관이 존재한다면 이는 지표변수들이 서로 구별되지 않고 중복된다는 의미가 되어 오히려 우려할 사항이 된다.

그 이유는 상관이 높은 두 지표변수는 인덱스를 구성할 때 서로 구별되는 독자적인 효과를 기대하기 어렵고 두 지표변수의 의미가 중복되었다고 볼 수 있기 때문이다. 따라서 형성적 측정모형에서는 지표변수들은 각각 독립적인 중요한 의미를 갖는 측면을 나타내므로 각 지표변수들이 상호 구별되는(discriminant) 관계를 기대할 수 있다.

#### 4.5.2 지표변수의 선택, 제거 기준 및 상호교환성(exchangeable)

반영적 측정모형을 적용하는 경우 고전적으로 내적 일관성이 있는, 즉 지표변수 간 상관성이 높은 지표변수를 선택함과 동시에 다른 지표변수들과 낮은 상관을 보이는 지표변수는 제거하여 공통성을 증가시키는 방향으로 지표변수의 취사선택을 실행한다. 반영적 지표변수들은 동질적인 모집단에서 추출된 것으로 볼 수 있으므로 반영적 지표변수들 간에는 상호교환성이 성립하며 특정 지표변수를 제거하거나 추가한다 하더라도 반영적 구성개념에는 거의 영향을 미치지 않는다. 따라서 충분한 수의 반영적 지표변수가 남아있다면 몇 개의 지표변수를 제거한다고 하더라도 그 영향은 다소간의 신뢰도 감소에 그칠 것으로 기대할 수 있다.

그러나 형성적 측정모형에서는 지표변수의 선택과 제거에 있어 판단 기준이 다르다. 형성적 측정모형에서는 지표변수의 선택에 있어 가능한 모든 독립적인 주요한 측면을 포괄하여 형성적 구성개념을 적절하게 나타낼 수 있도록 하는데 가장 중점을 둔다. 만약 형성적 지표변수의 선정 시 상관성이 높은 두 변수를 고른다면 이는 중복되는 측면을 나타내어 구성개념을 형성하는데 거의 도움이 되지 않을 것이다. 따라서 형성적 지표변수 선별과 제거의 기준으로 다른 지표변수와 상관성이 낮다는 것을 설정하지 않는다.

형성적 지표변수의 선별과 제거에는 매우 신중한 고려가 필요한데, 그 이유는 형성적 구성개념 전체에 미치는 영향 때문이다. 반영적 구성개념의 경우 동질적이라 여겨지는 일부 지표변수문항을 제거하거나 추가하더라도 구성개념에는 변화가 없으나, 형성적 모형에서는 하나의 지표변수를 제거하게 되면 그 지표변수가 대표하는 현상의 주요한 측면을 제외하는 것으로 나타나므로 형성적 구성개념 자체의 의미가 변화하게 된다. 형성적 구성개념의 독립적이고 주요한 측면을 나타내는 형성적 지표변수들 간의 상호교환성은 일반적으로 성립하지 않는다.

#### 4.5.3 모형 식별 (identifiability of model)

형성적 측정모형은 그 자체로는 개별적으로 식별되지 않고 다른 구성개념의 효과를 포괄하는 더 큰 모형에서 추정된다. 그러나 반영적 모형은 최소한 세 개의 효과 지표를 가지면 식별 가능하다(Simonetto, 2012)

#### 4.6 측정모형의 식별 오류 현황

반영적 측정모형과 형성적 측정모형은 기본 개념과 기본 가정 및 방법론 면에서 큰 차이가 있으나 기존의 많은 연구에서는 구성개념의 속성을 고려하고 반영적 측정모형과 형성적 측정모형 중 적합한 것을 비교 선택하는 과정이 없이 단지 반영적 측정모형만이 무차별적으로 도입되어 많은 식별오류(misspecification)의 경우가 나타났다고 한다(Jarvis, MacKenzie, & Podsakoff, 2003).

비교적 최근에 이르러 구성개념의 속성에 맞는 적절한 측정모형의 필요성이 강조되면서 기존 연구에 나타난 측정모형 적용에 대한 적합성 평가가 여러 영역에서 이루어졌고 특히 조직과 마케팅 분야에서 활발한 검증이 진행되었다. Roy 등(2005)은 Operations Management 분야의 주요 저널 4개를 선정하고 2002년부터 2006년 사이에 출판된 SEM 기반한 논문 총 94개를 선정하여 조사하였다. 94개의 저널에서는 642개의 구성개념이 발견되었으며 대부분의 일차측정모형(586개, 91.3%)에 반영적 모형이 적용되었고 단지 6개의 구성개념(0.9%)에만 형성적 모형이 사용된 것으로 나타났다. 이들 구성개념에 반영적 측정모형 및 형성적 측정모형이 올바르게 적용된 것인지를 평가하기 위하여 저자들이 공정한 기준을 마련하여 재평가한 결과, 반영적 일차측정모형이 적용된 구성개념 중 42.8%가 반영적 측정모형 보다는 형성적 측정모형이 더욱 적합한 것으로 나타났다.

또한 Jarvis, MacKenzie, and Podsakoff(2003)에 의하면 Journal of

Marketing 등 마케팅 분야의 상위 저널에서도 반영적 측정모형이 사용된 SEM모형 중 형성적 측정모형을 적용하는 것이 더 적합한 것으로 판단된 구성개념이 28%에 달하였다고 한다. 또한 Strategic Management 분야에서도 1994년부터 2003년 까지 10년 간 주요 저널에 257개의 구성개념이 출판된 바 있고 그 중 반영적 측정모형이 적용된 반영적 구성개념은 총 233개 이었다. 그 중 73개만이 올바르게 적용된 것으로 판단되었고 나머지 160개(68.7%)의 경우 반영적 측정모형이 아니라 형성적 측정모형이 적합한 것을 조사된 바 있다(Podsakoff, 2006).

## 5. 반영적 및 형성적 측정모형의 통계적 추정

구조방정식 모형 추정법은 크게 CB-SEM (Joreskog, 1978)과 PLS-SEM (Wold, 1982)의 두 가지 방향으로 발전하여 왔다. 이들 두 접근법은 구성개념과 지표변수 간의 관계를 추정하고자 하는 동일한 목적을 가지지만 통계적 개념에서 근본적으로 차이가 있고 특히 구성개념의 측정모형을 처리하는 방식에 큰 차이를 보인다.

### 5.1 CB-SEM 의 통계적 추정

CB-SEM은 요인분석법과 회귀분석법이 결합하여 확장한 것으로 볼 수 있다. 요인분석법의 원리는 먼저 지표변수의 분산을 공통분산(common variance)과 고유분산(unique variance)으로 나누는 것으로 즉 구성개념의 측정모형에서 다른 지표변수들과 공유하는 공통분산과 해당 변수에만 고유한 고유분산을 나누는 것이다. 고유분산은 다시 그 변수의 특수한 측면을 반영하는 규칙적이고(systematic) 신뢰도가 높은 특수분산(specific variance)과 랜덤하며 신뢰도가 낮은 오차분산으로 구분된다. 오차분산에는 자연과학에서 측정과정에서 발생하는 우연오차나 표본 추출 과정과 관련하여 발생하는 오차 등이 포함된다.

CB-SEM 은 추출된 공통분산이 구성개념을 표상하는 것이라고 간주하여 공통분산만을 이후 추정과정에 포함한다. 따라서 CB-SEM은 구성개념의 측정에 대하여 지표변수들의 분산은 하나의 구성개념과 개별적인 변수의 오차로 완벽하게 설명된다는 공통요인모형의 원리와 접근법을 따르고 있다. 공통요인모형 추정 접근법은 반영적 측정모형의 원리와 철학을 정확하게 구현하고 있다.

CB-SEM은 공통요인모형 접근법을 견지하면서도 또한 형성적 측정모형을 포함하여 추정할 수 있다. 이는 형성적 측정모형의 원인적 지표변수를 활용하는

방법으로 수행되는 바 이를 수행하기 위해서는 모형 식별을 위하여 부과되는 규칙을 따라야 하는데 이러한 규칙은 종종 CB-SEM의 이론적인 원리와 충돌하는 면이 있다(Hair, 2012; Bollen, 2015).

CB-SEM에서는 형성적 측정모형을 구현하는 하나의 방법으로 형성적 지표 변수와 반영적 지표변수를 함께 사용하는 합성적 모형(hybrid model)인 MIMIC 모형(Joreskog & Goldberger, 1975; Bollen & Bauldry, 2011; Henseler, 2017)의 형태로 형성적 측정모형을 구현하게 된다(그림 8). 그러나 MIMIC 방법은 모형식별의 문제를 해결하기는 하나 원인적 지표변수를 가지는 형성적 측정모형에서 구성개념의 분산이 저평가되는 오류는 해결할 수 없는데 그 이유는 공통요인 모형의 원리상 적절한 구성개념의 대리자(proxy)를 계산하는 과정이 상호 연관된 지표변수들을 필요로 하기 때문이다.

따라서 CB-SEM에서는 원인적 지표변수와 관련하여 형성적 측정모형을 단지 유사하게 구현할 뿐이라고 평가된다(Sarstedt, 2016). 또한 복합 지표를 사용하는 것을 고려하거나 다른 변수들과의 경로를 설정하는 등 여러 방법의 시도가 있었으나, 몇 연구자들은 제안된 방법들이 만족할 만한 해법이 아니라고 평가하고 CB-SEM은 형성적 측정모형을 추정하는데 적합하지 않다고 주장하였다(Hair, 2012; Peng & Lia, 2012).

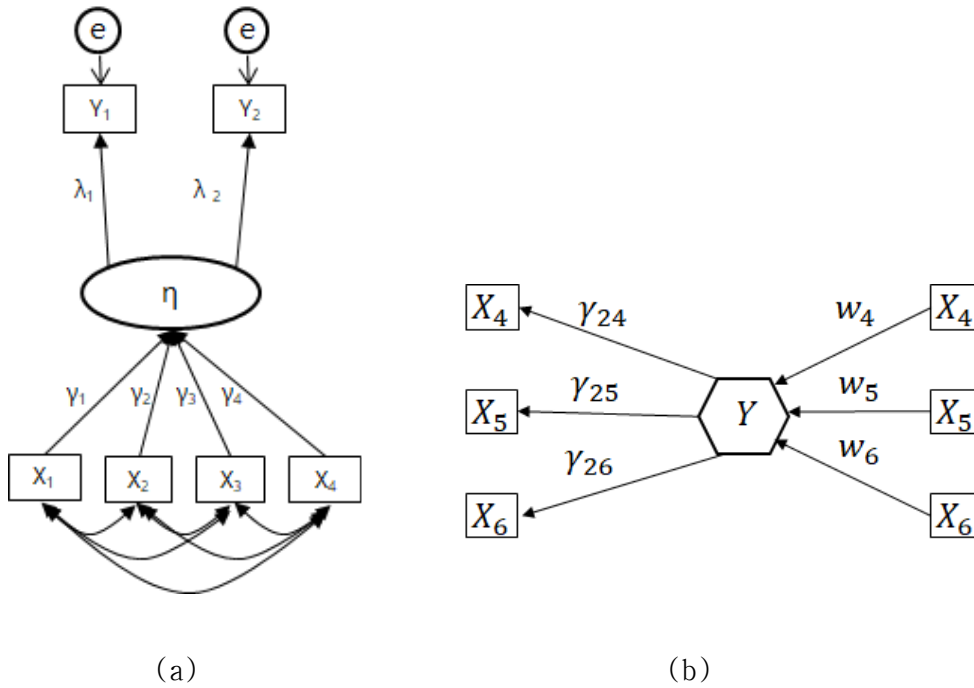


그림 8. Multiple Indicators - multiple causes (MIMIC) 모형: a) CB-SEM 에서 형성적 측정모형의 구현방법, b) PLS-SEM에서 반영적 측정모형의 구현 방법

## 5.2 PLS-SEM의 통계적 추정

PLS-SEM은 주성분분석법(Principal Factor Analysis, PCA)과 회귀분석법이 결합하여 구성개념의 측정모형과 구조모형을 추정하는 기법으로 확장된 것이다. Partial Least Square (PLS) regression<sup>11)</sup>은 주성분 분석과 회귀분석법이 단순 결합한 형태로 이를 모태로 구조방정식 모형을 추정하는 기법인 PLS-SEM으로 확

11) Partial Least Square (PLS) regression 기법은 산업응용 분야에서 빈번하게 활용되는 자료축소(data reduction) 기법의 하나로서 다중공선성(multicollinearity)이 큰 다수의 독립변수를 이용하여 종속변수를 예측하고자 하는 경우에 주로 사용한다. 자료축소를 목적으로 주성분 분석의 확장으로 주성분 회귀분석에서는 독립변수들의 변량을 최대한 포괄하여 데이터 축소를 시행하나 PLS 회귀분석에서는 독립변수들과 종속변수간의 변량을 최대한 포괄하여 그들 간의 관계를 가장 잘 설명하는 방향으로 데이터 축소를 수행한다. 만약 추출된 요인의 개수가 충분히 크다면 다변량 회귀분석과 동등하다(equivalent) (Tobias, 1995).

장되었다. PLS-SEM은 지표변수의 분산을 나누지 않고 전체 변량을 가지고 계산한다는 면에서 지표변수들의 상관만을 분석에 고려하는 CB-SEM방법과 근본적으로 다르다.

PLS-SEM은 공통분산과 고유분산을 합친 전체 분산을 가지고 구성개념들 사이의 관계를 추정한다. 지표변수들의 선형결합인 복합지수가 구성개념을 종합적으로 나타내어 준다는 기본개념에서 출발하고 구성개념은 또한 이론변수의 타당한 대리자로 기능한다고 보는 것이다. PLS-SEM은 형성적 측정모형의 원리와 철학을 따라 이를 구성개념의 측정에 적용하는 복합지수 모형 접근법을 따른다. PLS-SEM의 복합지수 모형 접근법은 추정 대상 측정모형이 반영적 측정모형이거나 형성적 측정모형이거나 관계없이 모두 동일하게 적용되고 있으며 모형식별의 문제는 발생하지 않는다.

PLS-SEM은 주성분 접근법을 기반으로 하므로 지표변수의 정보를 최대한 반영하는 가중선형결합의 형태로 복합지수를 생성한 후 이를 이용하여 반영적 측정모형을 추정한다. 이러한 형태는 CB-SEM에서 형성적 측정모형을 구성할 때 사용하는 MIMIC 모형과 유사하다(그림8-b).이 모형은 주성분 분석법에서 반영적 측정모형을 분석하는 방법으로 적용된다.

PLS-SEM은 형성적 측정모형에 대하여 모든 지표변수를 복합 지표변수로 간주하므로 원인적 지표변수와 같이 구성개념의 측정오차를 허용하는 측정모형은 적용하지 않는다. 반영적 측정모형에 대해서도 복합지수 모형 접근법으로 추정하므로 그 추정방법은 공통요인 분석법(common factor analysis, CFA)과 동일하지 않다. PLS-SEM은 반영적 측정모형을 추정할 때 편향된(biased) 추정치를 초래하여 요인계수의 절대값을 과대추정하고 그 결과 잠재변수 간의 상관이나 경로계수는 과소 추정하는 편향이 나타난다. 이러한 편향은 각 구성개념에 대한 요인계수가 큰 값이고 지표변수의 수가 많다면 매우 낮은 수준으로 나타난다(Ridgon, 2012).

이러한 편향의 문제점을 보완하기 위하여 consistent PLS가 개발되었는데 이는 반영적 측정모형의 추정과정에서 발생하는 과대 추정된 요인계수와 과소 추정된 잠재변수 간의 상관을 수정하여 보다 정확한 값으로 산출하기 위한 기법을 적용하였다(Dijkstra & Schermelleh-Engel, 2014).

### 5.3 일반화 구조성분분석법

황 등(Hwang et al, 2010)은 제 3의 대안으로 일반화 구조성분분석법 (generalized structured component analysis, GSCA)를 제안하고 다양한 조건에서 simulation 연구를 시행한 결과, 반영적 측정모형에 의거한 구조모형이 올바르게 식별된 경우에는 CB-SEM이 좋은 수행을 보였으나 모형이 올바르게 못하게 식별된 경우에는 GSCA가 보다 나은 수행을 보였다고 보고한 바 있다.

## 6. 공통요인 접근법과 주성분 접근법의 수리적 비교

PLS-SEM에서는 형성적 측정모형에 대하여 모든 지표변수를 복합 지표변수로 간주하므로 원인적 지표변수와 같이 구성개념의 측정오차를 허용하는 측정모형은 적용하지 않는다. 또한 반영적 측정모형에 대해서도 복합지수 모형 접근법으로 추정하므로 그 추정방법은 CB-SEM과 동일하지 않다. 반영적 측정구조를 가진 자료에 대하여 PLS-SEM을 적용하여 반영적 측정모형을 추정할 때에는 편향된(biased) 추정치를 초래하여 요인계수의 절대값을 과대추정하고 그 결과 잠재변수 간의 상관은 과소 추정하는 경향이 나타난다고 한다(Widaman, 1993). 여기서는 상기한 두 방법과 관련된 가장 기본적 형태인 공통요인분석법과 주성분분석법을 적용하여 수리적으로 두 방법 간의 차이를 비교하고자 한다.

### 6.1 추정 과정의 수리적 비교

공통요인 접근법과 주성분 접근법의 가장 큰 차이는 측정오차의 고려 여부이다. 많은 기존연구에서는 다양한 조건에서 두 접근법의 추정결과 나타난 요인계수/loading)을 비교 분석한 결과, 유사한 결과를 보고하여 두 방법 중 어느 것을 선택하든지 큰 차이가 없다는 주장을 제시하기도 하였다(Velicer, 1977; Velicer, Peacock, & Jackson, 1982).

CFA에서는 다수  $n$ 개의 지표변수에서 공유하는 변량을 소수  $p$  개의 잠재하는 공통요인으로 표현하고 이를 회전하여 소수의 요인으로 추출하게 된다. 이 방법을 행렬식으로 표현하면 다음과 같다.

$$R - U^2 = F_f F_f' = P_f \Phi_f P_f' = R^*_{ff} \quad (\text{식 2.4})$$

$R$ 은 관찰 변수간의 상관행렬 ( $n \times n$ ),  $U^2$ 는 고유분산(unique variance)의 대각행렬( $n \times n$ ),  $F_f$ 는 CFA로 추출된 회전하지 않은 요인 행렬 ( $n \times p$ ),  $P_f$ 는 CFA로 추출된 회전한 요인 패턴 행렬 ( $n \times p$ ),  $\Phi_f$ 는 CFA로 추출된 요인 간 상관행렬 ( $p \times p$ ),  $R^*_{\cdot f}$ 는 CFA의 적용 후 추정된 결과를 이용하여 계산된 재생된 상관행렬이다( $n \times n$ ). 그리고  $F'_f$ 와  $P'_f$ 는 각각  $F_f$ 와  $P_f$ 의 전치행렬(transposed matrix)을 나타낸다(Widaman, 1993).

CFA의 적용 후 추정된 결과를 이용하여 계산된 재생된 상관  $R^*_{\cdot f}$ 의 대각 성분은 측정된 반영적 지표 변수의 공통성(communality)을 나타낸다. 반영적 구성개념이 지표변수 변이의 공통분, 즉 공통성으로 수량화되므로 측정모형의 추정에는 개별 변수의 고유한 특성에 의하여 나타나는 고유분산( $U^2$ )을 제거한 축소된 상관행렬(reduced correlation matrix)이 사용되는 점이 특징적이다.

주성분 접근법이 CFA와 가장 큰 차이를 보이는 지점은 고유분산( $U^2$ )을 0으로 간주하여 원상관행렬(original correlation matrix)이 모형 추정에 사용되는 점이다. 즉 PCA는 CFA 중 고유분산이 0인 특수한 경우로 볼 수 있다. 만약  $n$ 개의 변수가 관찰되었는데 변수 수 보다는 크기가 작은  $p$ 개의 차원으로 요약된다면 형성적 측정모형은 다음 식 2.5와 같이 표현할 수 있다.

$$R = F_c F'_c = P_c \Phi_c P'_c = R^*_{\cdot c} \quad (\text{식 2.5})$$

$R$ 은 관찰 변수간의 상관행렬 ( $n \times n$ ),  $F_c$ 는 PCA로 추출된 회전하지 않은 요인 행렬 ( $n \times p$ ),  $P_c$ 는 PCA로 추출된 회전한 요인 패턴 행렬 ( $n \times p$ ),  $\Phi_c$ 는 PCA로 추출된 요인 간 상관행렬 ( $p \times p$ ),  $R^*_{\cdot c}$ 는 PCA의 적용 후 추정된 결과를 이용하여 계산된 재생된 상관행렬이다( $n \times n$ ). 그리고  $F'_c$ 와  $P'_c$ 는 각각  $F_c$ 와  $P_c$ 의 전치행렬(transposed matrix)을 나타낸다(Widaman, 1993).

PCA의 적용 후 추정된 결과를 이용하여 계산된 재생된 상관  $R^*_{\cdot c}$ 의 대각

성분은 전체 분산에 대한 비율(proportion)을 나타낸다.

## 6.2 추정결과의 수리적 비교

### 6.2.1 요인계수와 상관의 비교

측정오차 고려 여부는 결국 CFA 적용 결과로 추정된 공통요인계수에 비하여 PCA 적용 결과로 추정된 주성분 요인계수가 크게 추정되고, 그 직접적인 결과로 구성개념 간의 상관은 작게 추정되는 결과로 나타난다. 이러한 현상을 수리적으로 표현하면 다음과 같다:

먼저  $m$  개의 지표변수를 가지는 하나의 구성개념을 가정하는 경우 CFA 또는 PCA 적용 결과 추정되는 요인계수( $\hat{\lambda}$ )는 다음과 같이 표현된다.

$$\hat{\lambda} = [\lambda^2 + (\Delta h^2/m)]^{1/2} \quad (\text{식 2.6})$$

여기서  $\Delta h^2$  는 공통성 참값과 추정된 공통성과의 차이를 나타낸다.

PCA에서는 대각성분을 1로 하는 상관행렬을 적용하기 때문에 추정되는 대각성분은 모집단의 공통성을 필연적으로 과대 추정하므로( $\Delta h^2 > 0$ ) 필연적으로 추정되는 요인계수( $\hat{\lambda}$ )는 모집단의 요인계수인  $\lambda$ 보다 더 큰 값으로 과대추정될 것이다.

PCA 적용 시 나타나는 요인계수의 과대추정 정도는 공통성 수준과 하나의 구성개념에 대한 지표변수의 수에 따라 달라진다. 공통성 수준이 증가할수록 1과 실제의 공통성과의 차이가 줄어들게 되어  $\Delta h^2$  값이 감소하고 그 결과 요인계수 참값( $\lambda$ )와 추정되는 요인계수( $\hat{\lambda}$ )의 차이가 감소하여 과대추정의 정도는 감소하는 경향이 있다(Widaman, 1993). 또한 지표변수의 수가 증가하면 대각성분

의 수에 비하여 비대각 성분의 수가 상대적으로 빠르게 증가하고 그 결과 전체 상관행렬에서 대각성분의 영향력이 감소하므로 결과적으로 요인계수 참값( $\lambda$ )와 추정되는 요인계수( $\hat{\lambda}$ )의 차이가 감소하여 역시 과대 추정의 정도는 감소하는 방향으로 나타난다(Widaman, 1993; Schneeweiss & Mathes, 1995).

다차원 구성개념의 경우 PCA에 의하여 산출된 차원 간 상관 역시 추정오류를 나타내게 된다. 식 2.4 와 식 2.5에서 나타난 바와 같이 만약 요인계수가 과대 추정된다면 요인 간 상관은 필연적으로 과소 추정될 것이다.

Widaman(1993)에 따르면 이러한 차이는 특히 요인 당 지표변수의 수가 작을수록, 지표변수 간의 공통성이 작을수록 커지는데, 요인의 수와는 무관하다고 보고하였다. 구체적으로는 요인 당 지표변수의 수가 3개이면서 요인계수가 0.4 또는 0.6인 경우와 요인 당 지표변수의 수가 6개이면서 요인계수가 0.4인 경우 요인계수 참값과 추정치의 차이가 0.1을 상회하였다. 그 이외의 경우, 즉 요인계수가 0.8이거나, 요인 당 지표변수의 수가 12개인 경우, 및 요인 당 지표변수의 수가 6개이면서 요인계수가 0.6인 경우에는 요인계수 참값과 추정치의 차이가 0.1 미만이었다.

### 6.2.2 구조계수 추정치의 비교

앞에서 살펴본 바와 같이 PCA 적용의 결과 CFA에 비하여 요인계수가 크게 추정되고 구성개념 간의 상관이 과소 추정되는 편향이 발생하게 된다. 과소 추정된 상관계수는 구조계수 추정치가 과소 추정되는 결과로 이어지게 된다.

## 7. PLS-SEM의 통계적 추정과 분석 결과 평가

### 7.1 PLS-SEM 모형 산출 과정

편회귀모형(partial regression model)들은 2단계 과정을 포함하는 PLS-SEM algorithm 의 반복절차에 의해 추정된다. 첫 번째 단계에서 구성개념의 복합지수(composite score)가 추정되고 두 번째 단계에서 외부 가중치와 요인계수에 대한 최종추정치가 산출되며 구조모형의 경로계수(path coefficients)와 함께 내생잠재 변수에 대한  $R^2$ 값이 추정된다.

모수를 추정함에 있어서 PLS-SEM은 개별 지표변수의 기여도를 나타내는 가중치를 산출하는 방법으로 Mode A 와 Mode B의 두 가지 mode를 구별하여 적용한다. Mode A는 지표변수와 구성개념의 이변량 상관관계로부터 상관 가중치를 산출하여 선형결합의 가중치로 적용하며, Mode B는 최소자승법(ordinary least squares)으로 추정한 회귀계수를 가중치로 이용하는 방법이다. 회귀가중치는 개별지표변수와 구성개념의 상관과 더불어 지표변수 간의 상관을 함께 고려한 결과로 산출된다는 점에서 이변량 상관관계 가중치와 구별된다.

Mode A 와 Mode B의 가중치를 구하는 방법은 다를지라도 본질적으로 PLS-SEM에서 복합지표를 가중선형결합을 통하여 산출하는 것에는 변함이 없다. PLS-SEM에서는 Mode A와 Mode B를 각각 반영적 측정모형과 형성적 측정모형을 추정하는데 기본적으로 할당하고 있으나 연구자가 원하는 방법을 선택하여 사용할 수 있다(Vinzi et al, 2010).

### 7.2 PLS-SEM 모형 추정 모수

PLS-SEM algorithm은 지표변수의 값을 표준화한 데이터를 사용하여 수행

한다. 따라서 모수를 포함한 모든 산출 값은 표준화된 값으로 나타난다. PLS-SEM과정을 통하여 추정되는 모수는 다음과 같다.

1) 경로계수(path coefficient): 구조모형에서 편회귀모형에 의하여 잠재변수들에 대한 경로계수를 추정한다.

2) 외부 요인계수 (outer loading): 반영적 측정모형에 사용되며 구성개념과 관련된 지표변수의 단순회귀분석 결과로 추정된다. 이 때 종속변수는 구성개념의 점수로서 지표변수의 가중합인 복합지수가 이용된다.

3) 외부 가중치 (outer weight): 형성적 측정모형의 경우 사용되며 종속변수가 되는 구성개념에 대한 지표변수의 편다중 회귀분석에 의해 추정된 회귀계수를 개별 지표변수의 가중치로 산출한다.

4) 잠재변수의 평균 추정치가 추정된다.

### 7.3 반영적 측정모형의 평가

표 3에 나타난 바와 같이 PLS-SEM모형을 이용하여 반영적 측정모형을 적합한 경우, 추정과 관련한 평가 사항은 다음과 같다.

#### 7.3.1 내적 일관성 신뢰도(internal consistency reliability)<sup>12)</sup>

Alpha 계수는 모든 지표변수들은 동일하게 신뢰할 수 있다는 가정에 기반하는 반면, PLS-SEM은 개별 신뢰도에 따라 지표변수들은 처리한다. 지표변수들

12) 내적일관성신뢰도(internal consistency reliability)는 검사를 구성하고 있는 문항간의 일관성을 측정한다. 문항 간의 상관관계가 높다면 문항들이 일관성 있게 하나의 구성개념을 측정하고 있다고 간주할 수 있다. 고전검사이론에서 가장 빈번히 사용되는 것은 Alpha 계수로서 다음과 같은 공식을 사

용한다. 
$$\alpha = \frac{n}{n-1} \left( 1 - \frac{\sum S_i^2}{\sum S_x^2} \right)$$

n : 문항 수  $s_i^2$  : 문항 점수 분산  $s_x^2$  : 총점의 분산

의 외부 요인계수가 다름을 고려하여 계산하는 복합신뢰도 (composite reliability-CR( $\rho_c$ )) 를 사용한다.

$$\rho_c = \frac{(\sum_i l_i)^2}{(\sum_i l_i)^2 + \sum_i var(e_i)} \quad (\text{식 2.7})$$

여기서  $l_i$  는 특정 구성개념의 지표변수에 대한 표준화 외부 요인계수 값이고,  $e_i$  는 변수  $i$  의 측정오차이다. CR은 대체로 Alpha값과 동일하게 해석한다. CR은 0과 1사이의 값 가지며 0.6~0.7이면 수용 가능한 것으로, 좀 더 보수적인 기준으로 본다면 0.7~0.9사이이면 수용가능하다.(Nunally & Bernstein, 1994). 그러나 0.9이상이면 모든 indicator들이 구분되지 않는 항목이라는 의미이므로 바람직하지 않다. 매우 높은 CR은 의미상으로 중복되는 불필요한 항목을 다수 사용하였다는 의미로 해석 할 수 있으며 구성개념의 내용타당성 면에서 부정적인 결과가 도출되고 오차항의 상관관계를 높이므로 사용이 곤란하다. CR값이 0.6 이하이면 내적 신뢰도 확보에 실패한 것으로 본다.

### 7.3.2 수렴타당도(convergent validity)<sup>13)</sup>

지표변수와 관련된 구성개념의 복합 점수(composite score)의 관계의 정도를 평가하는데 구성개념이 설명하는 각 지표변수의 분산이 50% 이상이 되어야 한다는 기준을 적용한다(Hulland, 1999). 구성개념이 설명하는 지표변수의 분산은 외부 요인계수 제곱의 평균인 평균추출분산(average variance extracted, AVE)으로 평가하며 공통성과 동일한 개념으로 간주한다.

13) 수렴타당도(convergent validity)란 동일한 특성의 구성개념이 수렴한다는 의미에서의 타당도이다. 동일한 특성을 서로 다른 방법으로 측정한 점수들 간의 상관이 높을 때 수렴타당도의 근거를 제공한다고 간주한다. 지표변수들의 전체 분산 중 공통분으로 추출된 분산의 비율이 높다면 수렴타당도가 성립한다고 볼 수 있다.

$$AVE = \frac{1}{k} \sum_{i=1}^k (l_i)^2 \quad (\text{식 2.8})$$

여기서  $l_i$  는 특정 구성개념의 지표변수에 대한 표준화 외부 요인계수 값이며  $k$ 는 관련 지표변수의 수이다.

AVE가 0.5 이상이면 수렴타당도 기준을 만족한다고 간주하며 이는 외부 요인계수 값 0.708에 해당한다( $\sqrt{0.5} = 0.708$ ). 외부 loading이 0.7 이하인 지표변수에 대해서는 제거를 고려할 수 있으나 0.4~0.7의 범위의 외부 요인계수를 가지는 지표변수의 경우에도 무조건적으로 제거하지는 않으며, 해당 지표변수 제거로 인해 구성개념의 내용 타당성에 미치는 영향을 평가하여 결론을 내린다. 낮은 외부 요인계수를 보이는 경우라 할지라도 내용타당성에 영향을 미치는 것으로 판단되면 해당 지표변수를 제거하지 않는다.

### 7.3.3 판별타당도(Discriminant Validity)<sup>14)</sup>

판별타당도는 하나의 구성개념이 다른 구성개념과 적절히 구분되는지를 평가한다. 여기에는 두 가지 방법이 사용된다.

첫째, 지표변수의 cross loading<sup>15)</sup>을 평가하는 것으로 해당 지표변수가 관계된 구성개념의 외부 요인계수는 다른 구성개념과의 cross loading에 비하여 커야 한다(Hair, Ringle, Sarstedt, 2011).

둘째, 앞의 방법에 비하여 좀 더 보수적인 방법으로 Fornell-Larker 기준을 적용하여 평가하는 방법이다. 이는 구성개념의 AVE값의 제곱근은 다른 구성개념 간의 상관관계보다 커야 한다는 원리에 의하여 평가한다(Fornell and Larker,

14) 판별타당도(discriminant validity)란 서로 다른 구성개념이 구별되어 판별 가능하다는 의미에서의 타당도이다. 다른 특성을 측정된 구성개념이 그들 간의 상관이 높지 않을 때 판별타당도의 근거를 제공한다고 간주한다.

15) cross loading이란 하나의 구성개념에 속한 지표변수가 다른 구성개념에 대하여 가지는 일종의 외부 요인계수이다.

1981; Farrell and Rudd, 2009). 대각선에 해당 구성개념의 AVE의 제곱근을 배치하고 대각선 외에는 서로 다른 구성개념의 상관계수를 배치하여 비교한다. 이때 대각선의 값이 크다면 판별타당성을 보인 것으로 판단한다.

#### 7.4 형성적 측정모형의 평가

형성적 측정모형의 경우, 실증분석에 앞서 내용타당성 확인이 철저하게 이루어져야 하고 실증분석에서는 수렴타당도와 다중공선성의 측면이 평가되어야 한다. 표 3에 나타난 바와 같이 PLS-SEM모형을 이용하여 형성적 측정모형을 적합한 경우, 추정과 관련한 평가 사항은 다음과 같다.

##### 7.4.1 내용타당도(content validity)

형성적 지표변수가 구성개념의 모든 측면을 정확히 나타낸다는 것 확인하는 것이 필수적이다. 형성적 구성개념의 경우 반드시 내용 명세화(content specification)과정을 통하여 구성개념의 모든 측면을 충분히 설명할 수 있는 지표 변수들로 구성되어 있는지를 확인하고 반드시 전문가의 의견을 참고하여 이론적 근거를 제시하여야 한다.

##### 7.4.2 수렴타당도

형성적 측정모형에서의 수렴타당도는 중복 분석(Redundancy analysis)의 방법을 통하여 평가할 수 있다(김장현 외, 2014). 이는 동일한 구성개념에 대해 형성적으로 측정된 구성개념과 반영적 방법으로 측정된 구성개념이 높은 상관관계를 갖는지 확인 하는 방법이다. 두 구성개념 간의 경로계수가  $>0.9$  또는  $>0.8$

이며 수정된  $R^2$  값이 0.81 또는 0.64보다 큰 경우 수렴타당도가 확보되었다고 본다(Chin, 1998). 형성적 구성개념에 대응하는 반영적 구성개념이 연구 설계 시 구체화 되어 자료수집 과정에 포함되어야 중복 분석의 수행이 가능하다. 보다 간편한 방법으로 반영적 지표 대신 단일 문항 구성개념과 비교할 수도 있다.

#### 7.4.3 다중공선성(multicollinearity)<sup>16)</sup>

상관성이 높을 것으로 기대되는 반영적 지표변수와는 달리 형성적 지표변수 간에는 상관성이 낮을 수도 높을 수도 있다. 만약 서로 상관이 큰 지표변수들이 포함되어 있다면 형성적 측정모형의 경우 다중공선성의 문제가 발생하여 방법론적으로 해석 기준 등의 면에서 많은 문제점이 발생할 수 있다. 다중공선성이 발생하면 표준오차가 증가되어 추정된 가중치가 0과 통계적으로 유의한 차이를 보이기 어렵게 된다. 또한 다중공선성이 큰 경우 가중치가 잘못 추정되거나 가중치의 부호가 원래의 구성개념과 지표변수 간의 상관과 반대로 추정되는 경우도 발생할 수 있다.

다중공선성의 평가는 분산팽창요인(variation inflation factor, VIF)과 공차(tolerance)<sup>17)</sup>를 이용하여 수행하며 VIF가 5 보다 크거나 공차가 0.2 보다 작은 경우 다중공선성이 발생한 것으로 본다(Hair, Ringle, Sarstedt, 2011). 20%의 공차는 한 변수의 분산의 80%가 다른 변수들에 의하여 설명됨을 의미한다. 만약 공선성이 매우 높으면 해당 지표변수 중 하나를 제거하거나 그 변수들의 새로운

---

16) 다중공선성(multicollinearity)이란 회귀분석에서 독립변수들 간에 강한 상관관계가 나타나는 것이다. 다중공선성이 존재한다면 회귀식의 결정계수 값이 크나 개별 독립변수의 회귀계수가 유의하지 않은 이상 현상을 관찰할 수 있다. 형성적 측정모형은 다중회귀분석의 원리에 의거하여 추정 되므로 다중공선성은 반드시 해결하여야 하는 문제가 된다.

17) 분산 팽창 요인(VIF)이란 여러 독립변수들 중 하나를 선택하여 종속변수로 하고 다른 독립변수들로 회귀 분석을 시행하였을 때 (1-결정계수)의 역수이다. 만약 다른 독립변수들이 공통으로 해당 독립변수의 분산의 90%를 설명한다면 결정계수는 0.9 이고 분산팽창요인은 10이 될 것이다. 분산 팽창 요인과 공차(tolerance)는 역수의 관계로서 분산팽창요인이 10 이라면 공차는 0.1이 된다. 분산팽창요인이 크거나 공차가 작으면 다중공선성이 존재한다고 볼 수 있다.

조합으로 새로운 지표변수를 생성하여 해결 할 수 있다.

#### 7.4.4 형성적 지표변수들의 유의성과 적합성 평가

형성적 지표변수들을 평가함에 있어 또 다른 중요한 기준은 적절성(relevance)를 의미하는 외부 가중치이다. 외부 가중치의 크기는 구성개념에 대한 형성적 지표변수들의 상대적 기여도를 의미한다. 형성적 측정모델의 외부 가중치의 크기는 반영적 측정모델의 외부 요인계수보다 일반적으로 작다. 형성적 지표변수가 구성개념에 실제로 기여하는지 여부는 외부 가중치의 통계적 유의성으로 평가할 수 있다. 부트스트랩법을 이용하여 외부 가중치가 0일 수 있는지 여부를 검정한다.

많은 수의 지표변수가 사용되더라도 외부 가중치의 통계적 유의성을 유지할 수 있는 지표변수의 수는 제한적이다. 만약 지표변수 간에 서로 상관관계가 전혀 없다면 외부 가중치의 최대값은  $1/\sqrt{n}$ 이다. 예를 들어 10개의 지표변수가 있다면 외부 가중치의 최대값은  $1/\sqrt{10}=0.316$  으로 제한된다. 따라서 통계적으로 유의하지 않은 외부 가중치를 보이는 지표변수라 하더라도 무조건 측정모형의 질이 낮은 것으로 판단해서는 안 되고 그 지표변수가 단독으로 구성개념에 미치는 '절대적 영향력'으로 판단해야 하며 이는 단순회귀분석을 통한 외부 요인계수로 평가한다. 이때 0.5 이상의 외부 요인계수 값을 보이면 제거하지 않는 것이 원칙이다. 만약 외부 가중치가 유의하지 않으면서 동시에 외부 요인계수 값이 0.5 이하이면 이론적 적절성과 개념적 중복성을 고려하여 제거 여부를 결정한다. 더불어 내용타당성의 관점에서 형성적 지표변수의 적절성을 확인하는 것도 매우 중요하다.

PLS-SEM은 정규분포를 가정하지 않으므로 모수적 방법으로 유의성 검정을 적용할 수 없어 비모수적 부트스트랩 절차를 거쳐 통계적 유의성을 확인할 수 있다. 일반적으로 5000개의 본 표본과 동일한 크기의 부트스트랩 표본을 이용하여

외부 가중치가 0이라는 귀무가설을 검정한다. 한 가지 고려할 점은 표본이 다양한 만큼 잠재변수 값의 부호가 양에서 음으로 또는 음에서 양의 값으로 변동될 수 있으며 이에 따라 외부 가중치의 추정값의 부호도 달라질 수 있으며 이를 부호의 불확정성(sign indeterminacy)이라고 한다.

부호 변이를 처리하는 방법으로는 다음의 세 가지 방법의 선택이 가능하다. 먼저 가장 보수적인 ‘no sign change’가 있다 두 번째로는 부트스트랩 표본에서 나온 결과를 원래 표본에서 나온 부호로 변경하는 ‘à individual sign change’가 있으며 가장 liberal 한 결과를 제공한다. 세 번째, 절충안으로 한정적인 몇 개의 부호를 변경하는 ‘à construct-level sign change’가 있다.

표 3에는 반영적 측정모형과 형성적 측정모형에 대하여 PLS-SEM모형을 적용하여 추정한 후 평가할 사항을 정리하였다.

표 3. PLS-SEM모형 추정과 관련한 평가 사항

평가단 계	구 분	평가 사항	기준
			CR 0.6~0.7이면 수용가능
측정모 형 평 가	반 영 적	내적 일관성 / 구성개념 신뢰성	AVE 0.5 이상. 외부 요인계수 값
		지표변수 신뢰성	이 0.7 이상
		수렴타당도 (convergent validity) 관별타당도 (discriminate validity)	AVE값의 제곱근은 다른 구성개념 간의 상관관계 초과
	형 성 적	수렴타당도 (convergent validity) 지표변수들 간의 다중공선성	중복 분석: 수정된 R <sup>2</sup> 값이 0.81, 0.64 초과
		유의성 및 외부 가중치의 타당성 (relevance)	VIF가 5 초과 / 공차가 0.2 미만 외부가중치 유의
구조모 형 평		결정계수 (R <sup>2</sup> ) 예측적합성(predictive relevance:	R <sup>2</sup> : small-0.02, medium- 0.13, large-0.26

	$Q^2$	$GoF = \sqrt{AVE\overline{R^2}}$ ,
가	경로계수의 크기와 유의성	small-0.1, medium-0.25,
	효과크기( $f^2$ , $q^2$ )	large-0.36

CR: Composite Reliability, AVE: Average Variance Extracted, VIF: Variance Inflation Factor, GoF: Goodness of Fit.

## 7.5 PLS-SEM 분석 결과

PLS-SEM 구조 모형 분석결과의 확인 시 다음과 같은 순서를 따라 점검한다. 1단계로 다중공선성 관련 사항을 확인한다. 2단계로 구조모형 경로계수의 유의성과 적합성을 확인한다. 경로계수는 표준화된 회귀계수이므로 상대적인 효과로 해석하며 필요하다면 직접효과 및 간접효과도 확인 한다. 3단계로 결정계수 값을 확인하고 수정된 결정계수를 다양한 표본에서 나온 결과와 비교할 때 사용한다.

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}}, \quad (\text{식 2.9})$$

$$Adjusted R^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1} = R^2 - (1 - R^2) \frac{n-1}{n-p-1} \quad (\text{식 2.10})$$

4단계에서는 효과크기를 구하여 상호 비교한다.

$$f^2 = \frac{R_{square_{with\ all\ variables}} - R_{square_{Excluded}}}{1 - R_{square_{with\ all\ variables}}} \quad (\text{식 2.11})$$

f-square는 모형에서 특정 외생 구성개념이 제거되었을 때와 비교하여 효과 크기를 구한다. 평가 기준은 0.02 as small, 0.15 as medium, and 0.35 large 이다.

5단계로는 Blindfolding 및 예측적합성(predictive value)  $Q^2$ 과 효과크기  $q^2$ 를 확인 한다.

CB-SEM 의 적합도 통계량은 관찰된 공분산 행렬과 모형에 의하여 재생된

공분산 행렬의 차이에 의하여 산출되는 반면, PLS-SEM에서 적합도는 종속변수의 관측치 또는 근사치의 분산을 모형에 의하여 예측된 값이 얼마나 설명하는 지에 의하여 평가되므로 적합도의 의미가 서로 다르다(Hair et al, 2012a).

## 7.6 2차 계층 구조의 추정과 평가

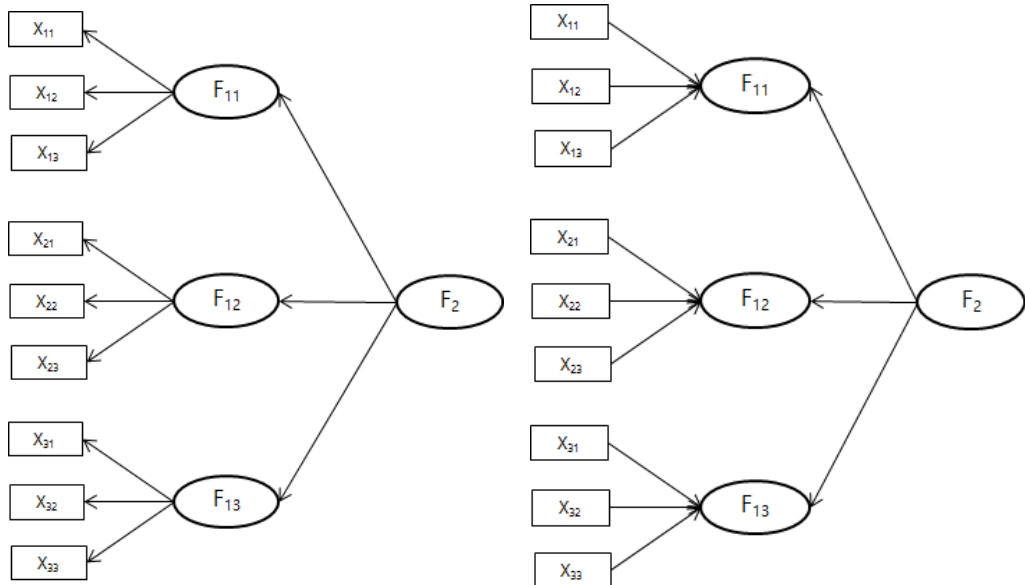
2차의 계층구조를 가진 모형의 경우 다음과 같은 순서로 모형을 추정한다 (Wrtzels et al, 2009).

1) 1차 잠재변수의 외부모형(outer model)을 수립한다.

2) 2차 잠재변수의 모형을 수립한다.

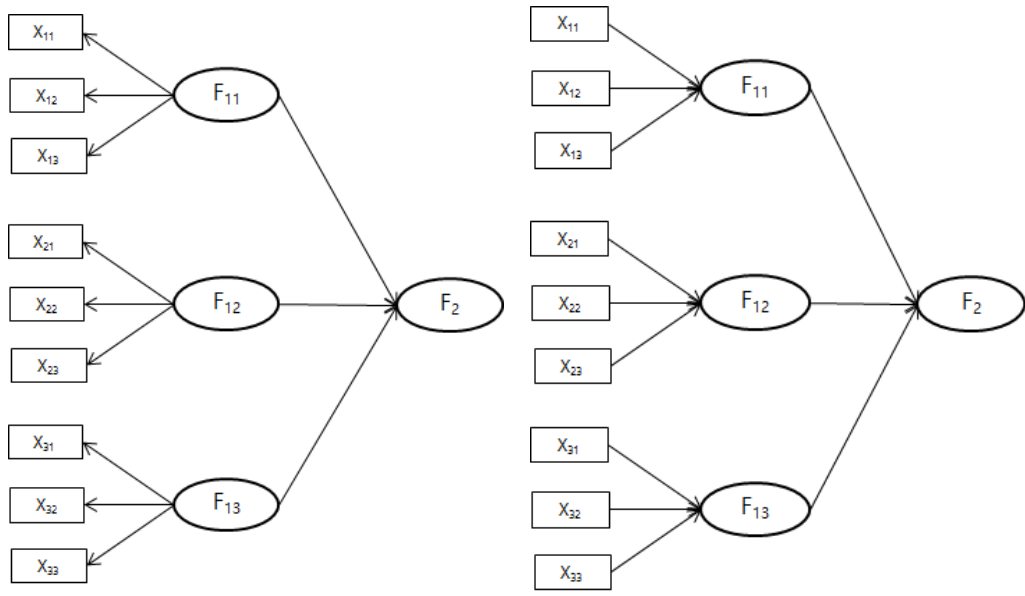
3) 계층구조 모형의 PLS path modeling을 통한 추정을 수행한다. 1차 요인 계수, 2차 요인 계수 및 구조모수 추정. 통계적 추론을 위한 표준오차와 t통계량 계산을 위하여 비모수적 부트스트랩 과정이 수행될 수 있으며, 잠재변수의 심리측정적 특성과 구조 관계를 평가한다.

1차와 2차 구성개념의 계층 구조를 가진 경우는 그림 과 같이 1차 2차 구조의 조합을 통하여 가능한 최대 4 개의 구조, 즉 1차와 2차가 모두 반영적인 구조, 1차는 반영적 구조이고 2차는 형성적인 구조, 1차는 형성적 구조이고 2차는 반영적인 구조, 1차와 2차가 모두 형성적인 구조 등으로 나타낼 수 있다(그림 9).



(a) Type 1

(b) Type 2



(c) Type 3

(d) Type 4

그림 9. 4 가지의 가능한 이차 측정모형 구조: (a) 1형 - 1차 및 2차 모두 반영적; (b) 2형 - 1차는 형성적 구조이고 2차는 반영적인 구조; (c) 3형 - 1차

는 반영적 구조이고 2차는 형성적인 구조; (d) 4형 - 1차와 2차가 모두 형성적인 구조(Becker et al, 2012)

PLS-SEM의 2차 계층모형의 추정에는 반복 지표변수 모형(Mode A 또는 Mode B), Hybrid 지표변수 모형(Mode A 또는 Mode B) 또는 2 단계 모형 등 세 가지의 서로 다른 접근법을 적용할 수 있다. Mode A는 지표변수와 구성개념의 이변량 상관관계로부터 상관 가중치를 산출하여 선형결합의 가중치로 적용하며, Mode B는 최소자승법(ordinary least squares)으로 추정한 회귀계수를 가중치로 이용하는 방법이다. 그림 10은 반복지표변수 접근법으로서 다수의 1차 구성개념에서 사용된 지표변수가 모두 2차 구성개념에 반복적으로 사용되어 모형추정에 이용되는 방법이며, 그림 11은 혼합 접근법으로서 전체 지표변수를 적절하게 나누어 절반은 1차 구성개념의 추정에 다른 절반은 2차 구성개념의 추정에 사용하는 방법이다. 2단계접근법은 먼저 1차 구성개념을 중심을 추정을 시행하고 2차 구성개념은 1차 구성개념의 차원점수, 즉 주성분점수를 이용하여 2단계로 추정하는 방법이다.

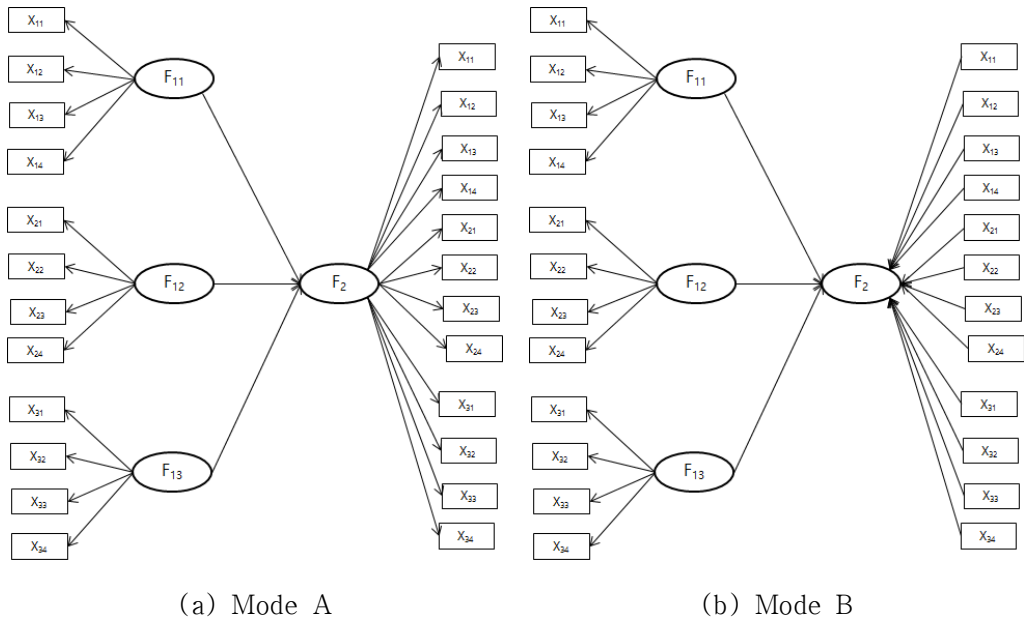


그림 10. PLS-SEM의 2차 계층모형의 추정을 위한 반복 지표변수 모형 (Repeated Indicator Model): (a) Mode A, (b) Mode B (Becker et al, 2012)

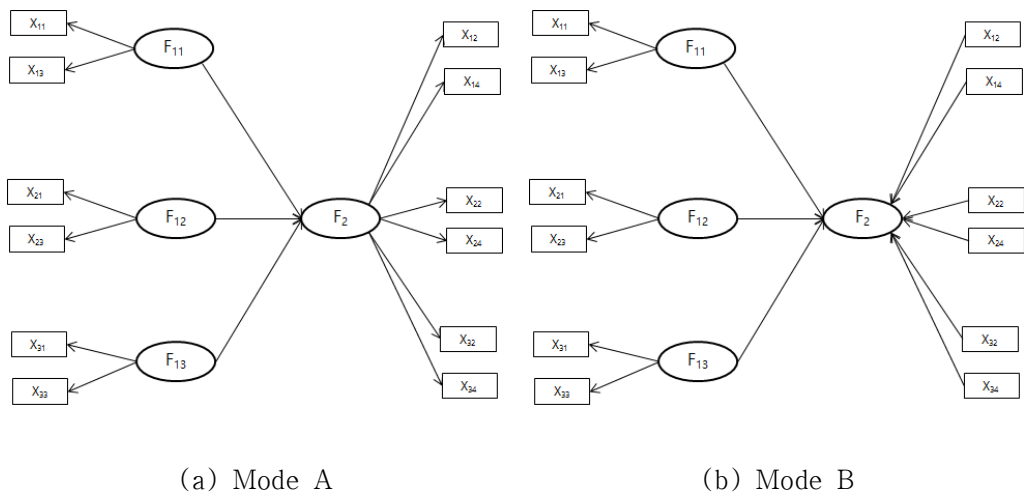


그림 11. PLS-SEM의 2차 계층모형의 추정을 위한 Hybrid 모형: (a) Mode A, (b) Mode B (Becker et al, 2012)

추정된 2차 모형의 평가 사항은 다음과 같다.

1) 반영적 1<sup>st</sup> 차 모형에서는 요인계수>0.7, AVE>0.5, CR, 0.6~0.9 등의 조건을 충족하여야 한다.(Nunally & Bernstein, 1994).

2) 공통성(communality)은 PLS-SEM모형에서는 AVE 값으로 나타나며 0.5를 기준으로 0.5를 상회하는 값을 갖는 것을 유효하다고 평가한다(Fornell and Larker, 1981).

3) R<sup>2</sup> 의 효과크기에 대하여 기준을 가지고 평가한다. small-0.02, medium-0.13, large- 0.26(Cohen, 1988)

4) global goodness of fit (GoF) =  $\sqrt{AVE \cdot R^2}$  (Amato et al, 2004)

GoF 값을 검토하여 small-0.1, medium-0.25, large-0.36 (Wetzels, 2009; Tenenhaus et al, 2005)로 간주한다.

### Ⅲ. 연구목적

이 연구는 다음과 같은 목적으로 수행되었다.

첫째, 다양한 건강 관련 삶의 질 측정도구 (SF-36, EQ-5D, COHIP, OIDP)의 측정구조의 형성적 및 반영적 특성을 고찰하고 적절한 분석모형을 수립한다.

둘째, 반영적 측정모형에 의하여 생성된 데이터에 대한 CB-SEM과 PLS-SEM 방법을 적용한 분석결과를 비교 평가한다. 이를 통하여 반영적 측정모형과 형성적 측정모형을 포괄하는 자료에 PLS-SEM을 적용함이 합리적인지 여부를 검증한다.

셋째, 건강 관련 삶의 질 도구의 실증적 자료를 이용하여 PLS-SEM을 적용하여 분석하고 평가한다. 더불어 PLS-SEM 모형의 교차타당도를 검증한다.

넷째, 건강 관련 삶의 질 측정도구에서 요약점수를 얻는 기존의 다양한 방법과 PLS-SEM 방법을 적용하여 그 결과로 나온 요약점수를 비교한다.

## IV. 연구방법

### 1. 다양한 건강관련 삶의 질 측정도구의 특성 고찰

#### 1) 건강관련 삶의 질(HRQoL) 도구

다양한 HRQoL 도구 중 반영적 측정구조를 가지는 SF-36 과 형성적 측정구조를 가지는 EQ-5D를 선정하여 세부적 지표변수 및 구성에 대하여 고찰하고 적절한 분석모형을 제시하고자 하였다.

#### 2) 구강건강관련 삶의 질(OHRQoL) 도구

다양한 OHRQoL 도구 중 반영적 측정구조를 가지는 COHIP와 형성적 측정구조를 가지는 COIDP를 선정하여 세부적 지표변수 및 구성에 대하여 고찰하고 적절한 분석모형을 제시하고자 하였다.

### 2. 데이터 모의실험 연구

요인계수의 크기, 지표문항의 수, 표본 크기 등의 조건을 달리하여 반영적 측정구조를 가지는 데이터를 반복 생성하고 이를 CB-SEM과 PLS-SEM 의 방법으로 각각 분석한 결과를 비교하여 PLS-SEM과 관련된 비일관성(inconsistency)의 크기와 방법을 평가하였다.

#### 1) 데이터 생성

다양한 크기의 요인계수(0.4, 0.6, 0.8), 3종의 차원 당 문항 수(3, 6, 12 개) 및 표본 크기(150, 300, 500, 1000)를 조합하여 각 조건에 맞는 데이터셋을 1000개 생성한다. 데이터는 모두 반영적 측정구조를 기반으로 하여 생성하였다. 이를 위하여 MPlus 버전 7.3(Muthen & Muthen, Los Angeles, CA, USA) 통계 프로그램을 사용하였다.

## 2) 데이터 분석

먼저 CB-SEM 모형을 사용한 분석을 하여 1000개 표본에서 추정된 요인계수의 평균, 참값과의 차이 및 평균제곱오차 등을 계산하였다. MPlus 버전 7.3(Muthen & Muthen, Los Angeles, CA, USA)을 이용하여 이 과정을 수행하였다.

PLS-SEM 모형을 사용한 분석에는 주성분 분석을 적용하기 위하여 SAS 버전 9.3 (SAS Inc., Cary, NC, USA)이 활용되었고 동일하게 1000개 표본에서 추정된 요인계수의 평균, 참값과의 차이 및 평균제곱오차 등을 계산하였다.

## 3. 건강관련 삶의 질 측정 자료에 대한 PLS-SEM 분석

### 1) PLS-SEM 모형 적용

반영적 측정구조를 가지는 SF-36, 형성적 측정구조를 가지는 EQ-5D, 복합적 측정구조를 가지는 COHIP 및 형성적 측정구조를 가지는 COIDP에 대하여 PLS-SEM모형을 적용한다.

### 2) 교차타당도(cross-validation) 검증

확보한 건강관련 삶의 질 실증자료 중 표본이 크고 형성적 및 반영적 측정 구조가 동시에 포함된 COHIP 자료를 사용하여 교차타당도를 검증한다. 자료를 반으로 나누어 각각을 PLS-SEM 모형으로 분석하여 외부 가중치, 외부 요인계수 및 구조계수 등의 크기를 비교한다.

#### 4. 현행 방법과 PLS-SEM 적용에 의한 통합점수의 비교

현행 건강 관련 삶의 질 측정 자료 통합점수 요약 방법인 평균점수 및 합산점수 이용법과 PLS-SEM 모형을 이용한 방법에 의한 가중치를 적용하여 구성 개념 점수를 계산하고 결과를 비교 분석한다.

## V. 연구결과

### 1. 건강관련 삶의 질 측정도구 고찰

#### 1.1 36-item Short Form Health Survey(SF-36)

##### 1.1.1 SF-36

SF-36은 Medical Outcome Study(MOS)에서 건강 관련 삶의 질을 조사하기 위하여 만들어진 프로파일 로서 임상현장과 임상연구, 건강정책 평가, 일반 국민 건강조사 등에서 활용되기 위하여 고안되었다. SF-36은 여덟 개의 건강개념을 평가하는 다문항 조사도구이다.

여덟 개의 건강개념은 건강문제로 인한 신체적 활동의 제한, 신체적 문제로 인한 역할의 제한, 신체적 통증, 정신건강(심리적 어려움과 안녕), 정서적 문제로 인한 역할의 제한, 정서적 문제로 인한 사회적 기능의 제한, 활력(에너지와 피로) 및 전반적 건강 인식 등 이다. 여덟 개의 차원은 SF-36의 문항 중 35개의 문항으로 측정되며, 나머지 하나의 문항은 1년 전에 비하여 현재의 건강상태를 어떻게 인식하는지에 대하여 질문하고 있다.

##### 1.1.2 SF-36의 8개의 차원

#### 1) 건강문제로 인한 신체적 활동의 제한(Physical functioning)

이 건강개념을 측정하기 위하여 응답자에게 일상생활에서 신체적 활동이

제한되었는지를 10개의 항목으로 질문하고 3가지 수준(limited / limited a little /not limited at all) 중 하나로 응답하게 하였다. 구체적 항목은 달리기 등 격렬한 활동, 탁자 옮기기 등 중등도 활동, 쇼핑한 물건 옮기기, 한 번에 두 세 계단 오르기, 한 계단씩 오르기, 구부리기, 1마일 이상 걷기, 한 블록 걷기, 여러 블록 걷기, 혼자서 목욕이나 옷 입기의 10개 항목이다.

2) 신체적 문제로 인한 역할의 제한(Role limitation due to physical problems, Role-physical)

이 개념은 4개의 항목을 포괄하고 있다. 구체적 항목은 지난 4주간 신체적 문제로 인하여 일하는 시간을 줄여야 하였는지, 일을 덜하였는지, 일의 종류에 제한을 받았는지, 일을 수행하는데 어려움이 있었는지 등이며 응답자는 ‘예’와 ‘아니오’의 두 답지 중 하나를 선택한다.

3) 신체적 통증(Bodily pain)

신체적 통증은 두 문항으로 측정되었다. 얼마나 통증을 경험하였는지를 none/ very mild/ mild/ moderate/ severe/ very severe 의 6 수준의 답지로 질문하였고 또한 통증 때문에 정상적인 생활이 방해받은 정도를 not at all/ a little bit/ moderately/ quite a bit/ extremely 의 5 수준으로 응답하게 하였다.

4) 정신건강(Mental health) 또는 정서적 안녕(emotional well-being)

정신건강은 지난 4주간 신경질적이었는지, 의기소침했는지, 안정되고 평온했는지, 상심했는지, 행복했는지 등의 5개 문항에 대하여 all of the time/ most

of the time/ a good bit of the time/ some of the time/ a little of the time/ none of the time의 6개의 수준으로 응답하게 하여 평가하였다.

5) 정서적 문제로 인한 역할의 제한(Role limitation due to emotional problems, Role-emotional)

이 개념은 세 개의 문항으로 조사되었다. 구체적 항목은 지난 4주간 정서적 문제로 인하여 일하는 시간을 줄여야 하였는지, 일을 덜하였는지, 일을 하지 않거나 조심해서 하여야 했는지 등이며 응답자는 '예'와 '아니오'의 두 답지 중 하나를 선택한다.

6) 사회적 기능(Social functioning)

사회적 기능은 두 문항으로 측정되었다. 지난 4주간 신체적 또는 정서적 문제로 인하여 가족, 친구, 이웃 등과의 정상적인 사회적 활동이 어느 정도까지 방해 받았는지를 not at all/ slightly/ moderately/ quite a bit/ extremely의 5 수준으로 질문하는 문항과 이러한 방해가 얼마만큼의 시간 동안 일어났는지를 all the time/ most of the time/ some of the time/ a little of the time/ none of the time의 5 수준으로 질문하는 문항이었다.

7) 활력(Vitality or energy)

활력은 4개의 문항으로 측정되었다. 지난 4주간 원기 왕성하였는지, 에너지가 가득했는지, 에너지가 고갈되었는지, 피곤하였는지 등의 문항에 대하여 all of the time/ most of the time/ a good bit of the time/ some of the time/ a

little of the time/ none of the time 의 6개의 수준으로 응답하게 하여 평가하였다.

#### 8) 일반 건강 인식

일반건강인식은 5개의 문항으로 측정되었다. 일반적으로 자신의 건강이 excellent/ very good/ good/ fair/ poor 중 어디에 해당한다고 보는지 선택하는 하나의 문항이 포함되었고 또한 다른 사람에 비하여 병에 잘 걸리는지, 더 건강한 편인지, 앞으로 건강이 나빠질 것으로 보는지, 본인의 건강상태가 매우 좋다고 보는 지 등은 definitely true/ mostly true/ don't know/ mostly false/ definitely false 의 5개 수준의 답지 중에서 응답하는 4개 문항이 포함되었다.

이상 8개 차원에 관한 개요는 표 4에 나타나 있다.

표 4. SF-36 삶의 질 도구의 8개 차원 과 점수의 해석

Concepts	No. of items	No. of levels	Meaning of scores	
			Low	High
Physical functioning	10	21	Limited a lot in performing all physical activities	Performs all types of physical activities
Role limitation due to physical problems	4	5	Problems with work or other daily activities as a result of physical health	No problems with work or other daily activities as a result of physical health, past 4 weeks
Bodily pain	2	11	Very severe and extremely limiting pain	No pain or limitations due to pain, past 4 weeks
Mental health (or emotional well-being)	5	26	Feeling of nervousness and depression all of the time	Feels peaceful, happy, and calm all of the time, past 4 weeks
Role	3	4	Problems with work or	No problems with work or

limitation due to emotional problems			other daily activities as a result of emotional problems	other daily activities as a result of emotional problems, past 4 weeks
Social functioning	2	9	Extreme and frequent interference with normal social activities due to physical and emotional problems	Performs normal social activities due to physical or emotional problems, past 4 weeks
Vitality	4	21	Feels tired and worn out all of the time	Feels full of pep and energy all of the time, past 4 weeks
General health perceptions	5	21	Believes personal health is poor and likely to get worse	Believes personal health is excellent

(Ware et al, 1992)

### 1.1.3 SF-36의 문항 구성과 차원 점수

SF-36의 문항 구성과 점수 부여체계는 다음 표 5와 같다. 각 문항에 대한 응답을 100점 기준으로 표 5에 나타난 바와 같이 환산한 후 8개 차원 마다 차원 별 문항의 평균을 구한다. 8개의 차원 점수는 건강 수준의 측정 및 건강상태의 변화 평가 등에 사용된다.

표 5. SF-36의 문항, 점수체계 및 차원

No.	Item	Response scoring	Domain
1	In general would you say your health is	1-Excel(100) 2-(75) 5-Poor(0)	GH1
2	Compared your age, general health	1-much better(100) 2-(75) 5-much worse(0)	---
3	Vigorous activities	1-Yes(0) limited 2-(50) 3-No not limited(100)	PF1

4	Moderate activities	1-Yes(0) limited 3-No not limited(100)	PF2
5	Lifting or carrying groceries	1-Yes(0) limited 3-No not limited(100)	PF3
6	Climbing several flights of stairs	1-Yes(0) limited 3-No not limited(100)	PF4
7	Climbing one flight of stairs	1-Yes(0) limited 3-No not limited(100)	PF5
8	Bending, kneeling, or stooping	1-Yes(0) limited 3-No not limited(100)	PF6
9	Walking more than a mile	1-Yes(0) limited 3-No not limited(100)	PF7
10	Walking several blocks	1-Yes(0) limited 3-No not limited(100)	PF8
11	Walking one block	1-Yes(0) limited 3-No not limited(100)	PF9
12	Bathing or dressing yourself	1-Yes(0) limited 3-No not limited(100)	PF10
13	Cut down the amount of time you spent on work (physical problem)	1-Yes(0) 2-No(100)	RP1
14	Accomplished less than you would like (physical problem)	1-Yes(0) 2-No(100)	RP2
15	Were limited in the kind of work (physical problem)	1-Yes(0) 2-No(100)	RP3
16	Had difficulty performing the work or activities (physical problem)	1-Yes(0) 2-No(100)	RP4
17	Cut down the amount of time you spent on work (emotional problem)	1-Yes(0) 2-No(100)	RE1
18	Accomplished less than you would like (emotional problem)	1-Yes(0) 2-No(100)	RE2
19	Didn't do work or activities as carefully (emotional problem)	1-Yes(0) 2-No(100)	RE3
20	To what extent, interfere social activities	1-not at all(100) 2-(75) 5-Extremely(0)	SF1
21	How much bodily pain	1-None(100) 2-(80) 6-Very scarce(0)	BP1
22	Pain interfere with normal work	1-not at all(100) 2-(75) 5-Extremely(0)	BP2
23	Feel full of pep	1-All(100) 2-(80) 6-None(0)	VT1
24	Been a very nervous person	1-All(0) 2-(20) 6-None(100)	MH1
25	Felt so down	1-All(0) 2-(20) 6-None(100)	MH2
26	Felt calm and peaceful	1-All(100) 2-(80) 6-None(0)	MH3

27	Have a lot of energy	1-All(100) 2-(80) 6-None(0)	VT2
28	Felt downhearted	1-All(0) 2-(20) 6-None(100)	MH4
29	Feel worn out	1-All(0) 2-(20) 6-None(100)	VT3
30	Been happy	1-All(100) 2-(80) 6-None(0)	MH5
31	Feel tired	1-All(0) 2-(20) 6-None(100)	VT4
32	How much of time, interfere social activities	1-All(0) 2-(25) 5-None(100)	SF2
33	Get sick easier	1-Def. true(0) 2-(25) 5-Def. false(100)	GH2
34	Healthy as anybody	1-Def. true(100) 2-(75) 5-Def. false(0)	GH3
35	Expect my health to get worse	1-Def. true(0) 2-(25) 5-Def. false(100)	GH4
36	My health is excellent	1-Def. true(100) 2-(75) 5-Def. false(0)	GH5

약어 설명 - PF:Physical Functioning, RP: Role limitation due to physical problems, BP: Bodily pain, GH: General health perceptions, SF: Social functioning, VT: Vitality, RE: Role limitation due to emotional problems, and MH: Mental health. (RAND Home page: Accessible at [https://www.rand.org/health/surveys\\_tools/mos/36-item-short-form/scoring.html](https://www.rand.org/health/surveys_tools/mos/36-item-short-form/scoring.html))

#### 1.1.4 SF-36의 신체성분점수(physical component score, PCS) 및 정신성분점수(mental component score, MCS)

여덟 개의 차원 점수는 신체성분점수와 정신성분점수로 묶어서 평가하기도 한다. 표 6에는 SF-36의 차원에 대한 주성분 분석 결과로 나타난 신체주성분 및 정신주성분과 개별 차원과의 상관 정도 및 상관계수를 표기하였다. Physical Functioning, Role Physical, Bodily Pain, General Health는 신체성분에 큰 요인 계수를 보이며 Mental Health, Role Emotional, Social Functioning은 정신성분에 큰 요인계수를 보이나 Vitality는 양 쪽 주성분에 걸쳐 중등도 이상의 연관을 보이고 있다(McHorney et al, 1993). 편의상 Physical Functioning, Role Physical, Bodily Pain, General Health는 신체성분으로 나머지는 정신 성분으로

분류하기도 한다.

표 6. SF-36의 주성분 분석 결과로 나타난 신체 및 정신주성분

Domain	Hypothesized association		Rotated Principal Component <sup>a</sup>	
	Physical	Mental	Physical	Mental
Physical functioning	+	-	0.88	0.04
Role-physical	+	-	0.78	0.30
Bodily pain	+	-	0.77	0.24
Mental health	-	+	0.12	0.90
Role-emotional	-	+	0.19	0.81
Social functioning	*	+	0.44	0.71
Vitality	*	*	0.59	0.57
General health perceptions	*	*	0.68	0.32

a: Correlation between each domain score and rotated principal component

+: Strong association ( $r \geq 0.7$ )

\*: Moderate to substantial association ( $0.3 < r < 0.7$ )

-: Weak association ( $r \leq 0.3$ ) (McHorney et al, 1993)

#### 1.1.5 SF-36 전체 점수

SF-36 도구의 각 차원을 통합하여 전체 점수를 구하는 방법은 공식적으로 제시되지 않았다. 관습적으로 연구자들은 8개 차원 점수를 평균하거나 전체 문항의 평균을 SF-36 전체 점수로 간주하여 사용하고 있다(차보경, 2016). Braizer 등(2002)은 SF-36의 축소판인 Short form-6D에 대하여 선호도 기반 측정치를 사용하여 여러 차원을 통합한 점수를 도출하는 방안을 제안하였다.

SF-36의 문항 모두를 이용하여 건강상태를 분류하기에는 너무 많은 경우의 수가 있어 처리가 곤란하므로 Role-physical과 General health perceptions를 제외한 6개의 차원을 대표하는 6 문항으로 이루어진 Short form-6D를 이용하였

다(표 7). 각 문항의 응답 수준은 각각 6, 4, 5, 6, 5, 및 5개 수준으로 이루어져 있어서 가능한 경우의 수는 총 18,000 가지이다. 가능한 경우를 예를 들면 ‘111111’ 은 모든 6개 차원에서 최상의 상태를 나타내고 ‘645655’는 최악의 상태를 나타내는 것 등이다. Braizer 등(2002)은 총 18,000 가지의 건강상태 중 49개의 건강상태를 선정하여 인구집단을 대표하는 표본에 대하여 선호도 조사를 실시하여 각 수준의 가치를 양적으로 평가하고 이를 모형화 하여 개개 차원의 수준별 가중치를 구하여 모든 수준의 건강상태에 해당하는 가치 평가를 수행하였다.

표 7. Short Form-6D (Braizer 등, 2002)

Level	Physical Functioning	Role limitation	Social Functioning	Pain	Mental Health	Vitality
1	vigorous activities not limit	no problem	not limit	no pain	none	all
2	Limit vigorous activities	limited by physical problems	limit a little of the time	pain but not interfere	feel tense a little	most
3	Limit a little moderate activities	accomplish less by emotional problems	limit some of the time	pain interfere a little	feel tense some	some
4	Limit a lot moderate activities	2+3	limit most of the time	pain interfere moderately	feel tense most of the time	a little
5	Limit a little in bathing		limit all of the time	pain interfere quite a bit	feel tense all of the time	none
6	Limit a lot in bathing			pain interfere extremely		

### 1.1.6 SF-36의 기존 분석 모형

SF-36이 심리측정의 원리에 의거한 도구이므로 기존의 많은 연구에서는 요인분석을 통하여 3차 구조방정식 모형을 적용하여 SF-36 도구의 구성타당도를 연구하고 있다(그림 12).

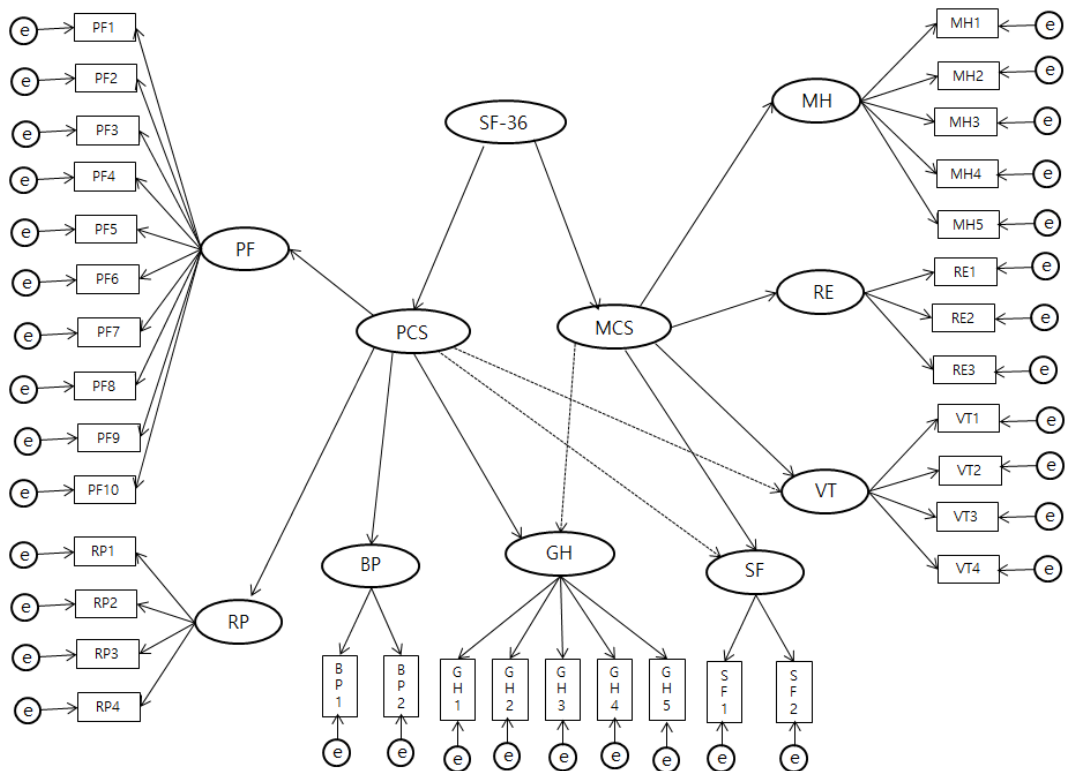


그림 12. 기존 연구에 나타난 반영적 3차 SF-36 구조모형(McHorney et al, 1993).

Solid line(————): Strong relationship, dotted line(-----): moderate relationship.

약어 설명 - SF-36: SF-36 score, PCS: Physical component score, MCS: Mental

component score, PF: Physical Functioning, RP: Role limitation due to physical problems, BP: Bodily pain, GH: General health perceptions, SF: Social functioning, VT: Vitality, RE: Role limitation due to emotional problems, and MH: Mental health.

한창완 등(2009)은 한국어판 SF-36을 이용한 연구를 수행한 바 있는데 35개의 문항으로 이루어진 8개의 1차 구성개념과 2개의 신체적 및 정신적 2차 구성개념, 또는 일반적 안녕(General Well-being) 까지 포함한 3개의 2차 구성개념 및 한 개의 3차 구성개념을 반영적 관계로 설정하여 구성타당도를 검증한 바 있다(그림 13과 그림 14). 그림 13과 그림 14를 종합적으로 고려하면 신체적 및 정신적 2차 구성개념 간의 상관이 0.935로 매우 크고 또한 3차 구성개념인 HRQoL과 세 개의 2차 구성개념 간의 구조계수가 0.85~0.935의 범위로 매우 크다.

이상과 살펴본 바와 같이 SF-36의 분석모형에 관한 기존 연구들은 모두 차원의 측정이나 차원 간의 관계를 나타내는 2차 또는 3차 구성개념의 관계에 있어서 반영적 관계를 적용하여 분석모형을 구성하고 있다. 그러나 각 차원들은 서로 구별되는 독특한 특성으로 구분되며 상호 교환 가능하지 않으므로 반영적 관계라고 볼 수 없다. 이는 판별타당도 등의 방법으로 심리측정적 방법에 의하여 규명된 바 있다(Gandek et al, 2004).

또한 그림 13과 그림 14에 나타난 바와 같이 구성개념들 간의 큰 상관과 구조계수를 고려한다면 2차 구성개념이 신체적 및 정신적으로, 또는 신체적, 정신적 및 일반건강상태로 구분되기 어려운 것으로 보이며, 하나의 상위 구성개념으로 통합하는 것이 보다 합리적이라 사료된다.

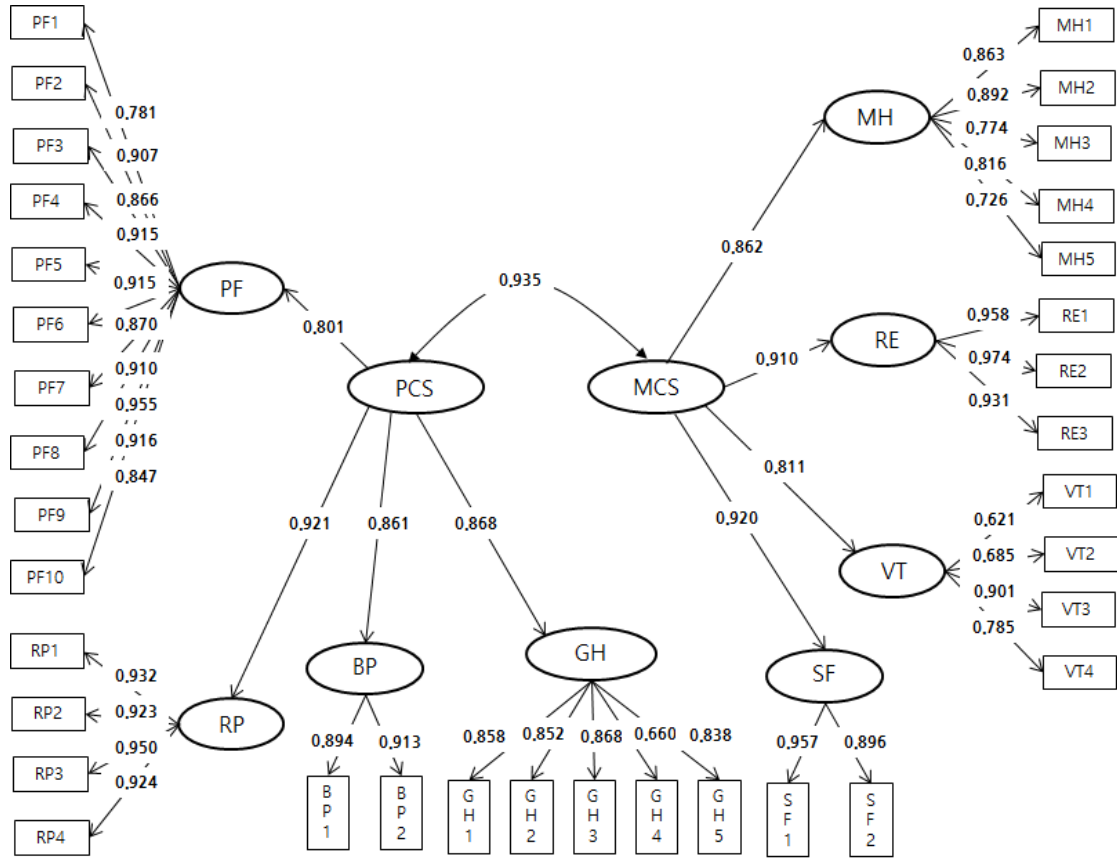


그림 13. 한국어판 SF-36을 이용하여 추정된 반영적 SF-36 구조모형: 2차 구성 개념 간 상관 모형 (한창완, 2009)

Standardized factor(or structural) coefficient 로 표시됨.

약어 설명 - PCS: Physical component score, MCS: Mental component score, PF: Physical Functioning, RP: Role limitation due to physical problems, BP: Bodily pain, GH: General health perceptions, SF: Social functioning, VT: Vitality, RE: Role limitation due to emotional problems, and MH: Mental health.

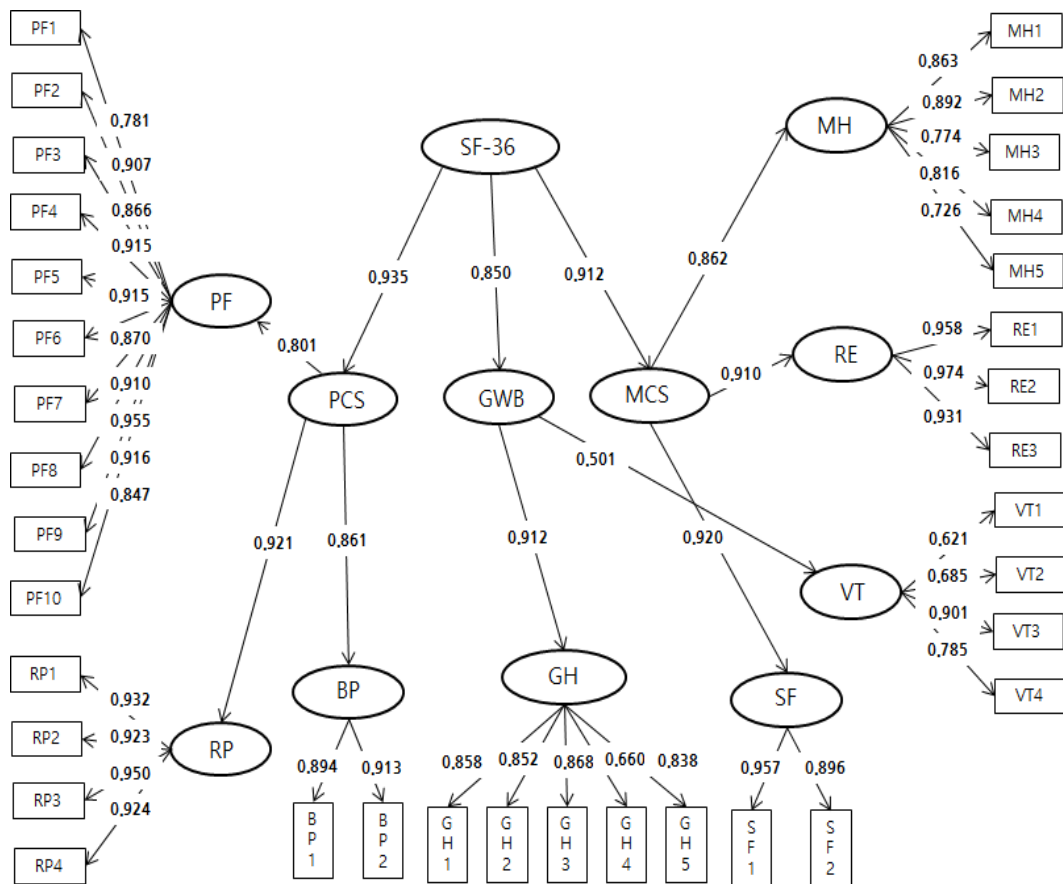


그림 14. 한국어판 SF-36을 이용하여 추정된 3차 반영적 SF-36 구조모형: 2차 구성개념 추가(한창완, 2009)

Standardized factor(or structural) coefficient 로 표시됨.

약어 설명 - SF-36: SF-36 score, PCS: Physical component score, GWB: General Well-Being, MCS: Mental component score, PF: Physical Functioning, RP: Role limitation due to physical problems, BP: Bodily pain, GH: General health perceptions, SF: Social functioning, VT: Vitality, RE: Role limitation due to emotional problems, and MH: Mental health.

### 1.1.7 본 연구를 위한 SF-36의 분석 모형

본 연구에서는 기존 연구 결과에 근거하여 8개의 1차 구성개념인 차원을 구분하였다. 그러나 기존 연구와 다르게 2차 구성개념의 관계는 형성적인 것으로 가정하였다. 그 이유는 각 차원은 서로 유사하거나 교환이 가능하지 않고, 각각이 구별되는 독특한 측면을 가지고 있으며, 각 차원에서의 측정수준이 원인이 되어 상위 구성개념의 변화를 일으키는 인과적 관계를 가정하는 것이 보다 적절하다고 사료되었기 때문이다. 또한 표 6(83쪽)과 그림 14 및 그림 15를 종합적으로 고려하여 신체적 성분과 정신적 성분이 명료하게 구별되지 않는다고 보았다. 따라서 본 연구에서는 1차 구성개념은 반영적 측정구조를 가지며 2차 구성개념은 형성적 특성을 가지는 2차 계층 구조방정식을 분석모형으로 선정하였다(그림 15).

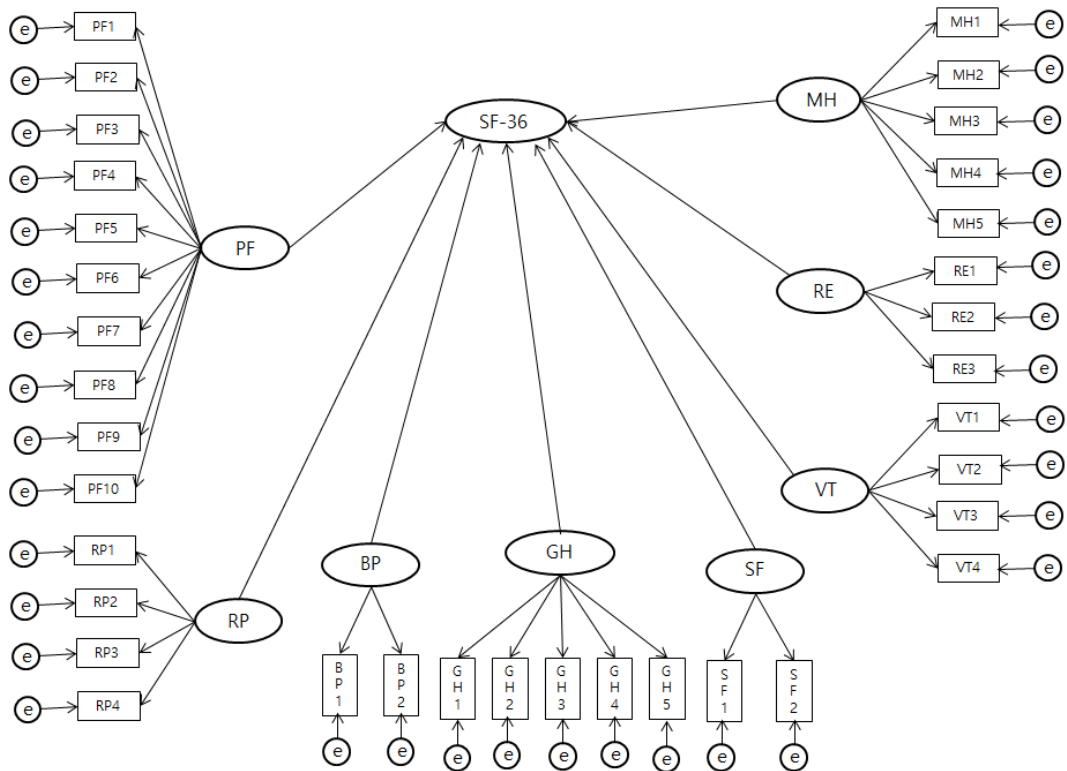


그림 15. 본 연구에서 가정하는 형성적 2차 구조를 가지는 SF-36 구조모형.  
약어 설명 - SF-36: SF-36 score, PF: Physical Functioning, RP: Role limitation due to physical problems, BP: Bodily pain, GH: General health perceptions, SF: Social functioning, VT: Vitality, RE: Role limitation due to emotional problems, and MH: Mental health.

## 1.2 EuroQol-5D (EQ-5D)

### 1.2.1 EQ-5D 개요

EuroQol 그룹에서 개발한 HRQoL 도구로서 세계적으로 가장 많이 사용되는 삶의 질 도구 중 하나이다. EQ-5D는 건강에 대한 전체론적(holistic) 관점에서 의학적 정의 뿐 아니라 독립적인 신체적, 정서적 및 사회적 기능을 포괄하는 건강의 개념에 기반 하여 만들어졌다. EQ-5D에 나타난 건강의 개념은 건강의 긍정적인 면(well-being) 및 부정적인 면(illness)을 포괄하고 있다(표 7).

EQ-5D는 짧으면서도 과학적 실험, 건강정책, pharmacoeconomics, 의료기관 등 여러 다른 장면에서 성공적으로 사용된 바 있다. EQ-5D는 설문지와 EQ visual analogue scale(EQ-VAS)로 구성되어 있다(그림 18). EQ-VAS는 응답자의 자가보고 전반적 건강을 VAS를 이용하여 표현하는 것이며 현재의 건강상태를 기록하고 또한 시간에 따른 변화를 살펴볼 수 있다. EQ-5D는 운동능력, 자기 관리, 일상활동, 통증/불편 및 불안/우울 등 5개의 차원에서 주관적 건강을 자가보고법을 사용하여 평가하는 것이다. 5개의 차원은 각각 3개의 장애 정도(심함, 보통, 없음)의 선택지가 있어서  $3^5=243$ 개의 건강상태의 합성이 가능하고 여기에 완전 건강과 죽음의 상태를 더하여 총 245개의 서로 다른 건강상태를 얻는다. 각각의 건강상태는 순위가 매겨지고 'utility'라고 부르는 하나의 점수로 변환된다. 'utility' 점수는 QALY의 한 표현이며 이는 비용효과분석에서 증거기반 결정을 수행하는 데 자주 사용된다(Gusi, 2010).

### 1.2.2 EQ-5D의 항목과 구성

EQ-5D HRQoL 측정도구는 표 7에 나타난 바와 같이 운동능력, 자기관리,

일상활동, 통증/불편 및 불안/우울에 관하여 오늘의 건강상태를 설문하는데 각각 하나의 항목이 할당되어 총 5개의 항목으로 구성되어 있고 응답자는 3개의 수준 혹은 5개의 수준으로 답변한다(표 8). 이 항목들은 각각 서로 다른 측면을 조사하고 있어 상호 교환이 불가능하고 각 항목에서의 측정수준이 원인이 되어 상위 구성개념의 변화를 일으키는 인과적 관계를 가정하는 것이 보다 적절하다고 사료되었기 때문에 그림 16과 같이 형성적 측정구조로 분석모형을 정리하였다.

표 8. 오늘의 건강상태를 질문하는 EQ-5D 도구

항목	선택지
운동능력	나는 걷는 데 지장이 없다
	나는 걷는 데 다소 지장이 있다
자기관리	나는 종일 누워있어야만 한다
	나는 목욕을 하거나 옷을 입는 데 지장이 없다
	나는 혼자 목욕을 하거나 옷을 입는 데 다소 지장이 있다
일상활동	나는 혼자 목욕을 하거나 옷을 입을 수가 없다
	나는 일상활동을 하는 데 지장이 없다
	나는 일상활동을 하는 데 다소 지장이 있다
	나는 일상활동을 할 수가 없다
통증/불편	나는 통증이나 불편감이 없다
	나는 다소 통증이나 불편감이 있다
	나는 매우 심한 통증이나 불편감이 있다
불안/우울	나는 불안하거나 우울하지 않다
	나는 다소 불안하거나 우울하다
	나는 매우 심하게 불안하거나 우울하다

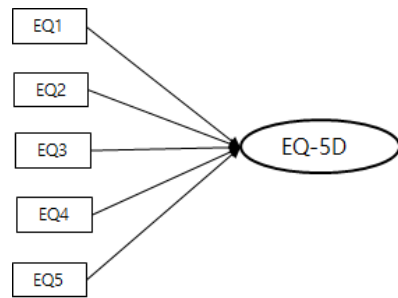


그림 16. EQ-5D의 개념적 모형

더불어 EQ-5D 측정도구에서는 Visual analogue scale 로 HRQoL 수준을 0에서 100의 범위에서 주관적으로 평가하게 하는 EQ-VAS 문항이 포함되기도 한다(그림 18).

문 2) 위에 그려진 눈금자에는 0에서 100까지의 숫자가 표시되어 있습니다.  
 100은 상상할 수 있는 최고의 건강 상태를 의미하고,  
 0은 상상할 수 있는 최저의 건강 상태를 의미합니다.  
 눈금자 상에 오늘 귀하의 건강 상태를 X로 표시해 주십시오.

이제, 눈금자 상에 표시하신 숫자를 아래 박스에 기입해 주십시오.

오늘의 건강 상태 =

21-23

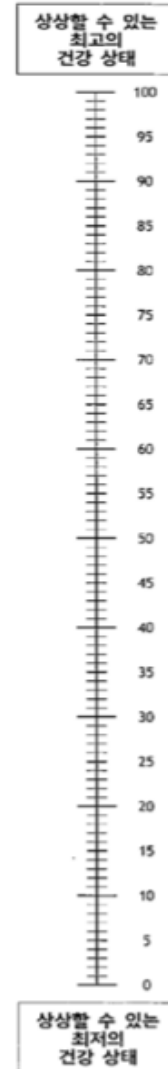


그림 17. EQ Visual Analogue Scale (EQ-VAS)

## 1.2.2 EQ-5D 점수 계산 방법

### 1) 선호도 가중치 사용법

EQ-5D는 죽음을 0으로 완전한 건강상태를 1로 정의하여 상대적인 좋고 나쁨을 단일한 수치로 표현함으로써 QALY 추산의 기반을 제공한다. EQ-5D에 의한 건강상태에 대한 가중치는 국가에 따라 차이가 있다는 것이 일반적인 견해로 받아들여지고 있어서 각 나라들에서 자신의 상황에 맞도록 가중치를 산출하고 있다.

EQ-5D의 선호도 가중치 추정법은 다음과 같은 순서로 진행한다.

먼저 조사 대상의 건강상태를 선택한다. EQ-5D는 운동능력, 자기관리, 일상활동, 통증/불편 및 불안/우울 등의 항목이 각각 3개 또는 5개의 수준으로 조사되며 각 항목의 수준을 연결된 다섯 자리의 수로 개별 건강수준을 표현하게 된다. 모든 문항에 1을 응답한 경우 11111로, 모든 문항에 3을 응답한 경우 33333으로 표현이 가능하며, EQ-5D에서 나타날 수 있는 모든 건강수준은  $243(3^5)$ 가지이다. 243가지의 건강 수준 모두를 응답자에게 답변하도록 하는 것은 현실적으로 무리이므로 기존 연구들에서는 그 중 42개에서 17개 사이의 건강수준을 선택하여 조사를 실시하였다.

다음 단계는 인구 집단을 대표하는 표본을 선정하여 시간교환법(Time Trade-Off, TTO)을 이용하여 건강상태의 가치 평가를 조사하는 것이다. TTO는 죽음보다 나은 가상의 건강상태에 대하여 그 상태로 x년을 사는 것과 완전한 건강상태로 t년을 사는 것이 동일한지를 물어서 TTO를 통한 해당 건강상태의 가치를  $t/x$  으로 추정하는 방법이다. 이때 완전한 건강상태의 가치는 최고값 1로 정의한다(그림 3, 13쪽). 이 TTO 점수는 어느 하나의 건강상태에 대하여 다른 건강상태에 대한 선호의 정도를 나타낸다.

세 번째 단계에서는 조사하지 않은 나머지 건강수준에 대해서 가치 예측을 한다. 일반적으로 조사된 건강상태의 가치 수준을 종속변수로 하고 다섯 항목의 해당 수준을 설명변수로 하여 회귀분석을 실시하여 다섯 항목의 각 수준의 회귀 계수를 얻고 이를 가중치로 하여 예측된 값을 건강상태의 가치 수준으로 간주하는 방법이 사용되고 있다. 개인의 가치평가에는 상호 차이가 있을 수 있으므로 이를 반영하여 고정효과모형(fixed effect model)이나 무선효과모형(random effect model) 등이 고려된다(강은정 등, 2006).

## 2) 합산점수 사용

EQ-5D의 다섯 문항의 수준을 좋은 건강상태가 큰 값을 가지도록 조정된 후에 그 합산점수를 HRQoL 점수로 이용하는 방법이다. 일반 연구자들이 많이 사용하는 방법으로 각 문항 또는 차원에 대한 가중치가 동일하다는 가정 하에서 사용한다고 볼 수 있다.

## 1.3 어린이 구강건강영향 프로파일(Child Oral Health Impact Profile, COHIP)

### 1.3.1 COHIP의 개요

COHIP은 Broder 등(2007a)이 건강관련 삶의 질의 개념적 모형에 의거하여 어린이의 구강건강 관련 삶의 질을 측정하는 도구로 개발하였다. 총 34개의 문항으로 개발되었고 이론적 모형에 따라 구강건강, 기능적 제한, 정서적 제한, 학교활동 제한 및 자신의 이미지 등 5개 차원으로 분류되었다. COHIP 설문은 지난 3개월 동안 구강 건강의 문제로 인하여 경험한 제한이 얼마나 빈번한지를 질문한다.

### 1.3.1 COHIP의 요인, 문항 및 구조 모형

경험적 데이터의 축적과 더불어 완결된 구조모형으로 COHIP의 측정모형을 확립하여 구성타당도를 검증하고자 하는 노력 중 Cho et al.(2016)에 의하여 4 요인 모형이 수정 보고되었다(표 9). 이는 원래의 이론적 5 요인 모형에서 기능의 제한과 학교활동 제한 간 상관이 0.92를 나타냄에 따라 이 두 요인을 합치고 더불어 오차분산이 큰 세 개의 지표변수를 제거하기로 결정하여 모형의 통계적 추정을 수행하였고 그 결과 그림 18과 같이 경험적 모형 추정 결과를 보고하였다(Cho et al, 2016). 기존연구에서는 심리측정적 방법론에 따라 CB-SEM을 적용하여 모든 관계가 반영적으로 정의되었다(Broder et al, 2007a; Broder et al, 2007b; Cho et al, 2016). 4개의 차원을 면밀히 살펴보면 'Oral Health'차원의 경우 지표변수들이 서로 교환 가능하지 않는 항목들이 드러난다. 가령 구강 내에 통증을 느끼는 것과 치아가 변색되는 것은 서로 유사한 특성이 아니고 구강건

강의 독특한 측면을 나타내므로 교환 가능하지 않다. 한편 다른 차원들은 대체로 구성개념의 특성에 따라 항목에 변화가 일어나는 반영적 관계를 보이는 것으로 간주될 수 있다. 여기에서도 네 차원 간의 관계는 서로 구별되는 측면을 나타내며 교환가능하지 않다. 따라서 이들 차원의 관계는 형성적 관계로 정의되어야 하는 것으로 사료된다.

표 9. 어린이 구강건강영향 프로파일(COHIP)의 요인 및 항목: CB-SEM의 확인적 요인분석에 의한 표준화 요인계수 (Cho et al, 2016)

Factor	Items	Factor loading
Oral Health (OH), 10 items	OH1 (pain)	0.55
	OH2 (mouth breathing)	0.44
	OH3 (discolored teeth)	0.41
	OH4 (crooked teeth)	0.36
	OH5 (sore spots)	0.49
	OH6 (bad breath)	0.47
	OH7 (bleeding gums)	0.48
	OH8 (food sticking)	0.52
	OH9 (cold sensitivity)	0.55
	OH10 (dry mouth)	0.48
Functional Well-being including school activities (Fu-S), 9 items	FU1 (trouble chewing)	0.57
	FU2 (difficulty eating)	0.61
	FU3 (trouble sleeping)	0.57
	FU4 (difficulty pronouncing)	0.55
	FU5 (difficulty being understood)	0.56
	FU6 (difficulty keeping clean)	0.49
	Sch2 (difficulty paying attention)	0.67
	Sch3 (not wanting to speak)	0.63
Social/Emotional Well-being (SE), 8 items	Sch4 (not wanting to go to school)	0.42
	SE1 (unhappy)	0.73
	SE2 (felt worried)	0.68
	SE3 (avoid smiling)	0.74
	SE4 (felt different)	0.68
	SE5 (worried)	0.71
SE6 (felt shy)	0.79	

	SE7 (teased)	0.52
	SE8 (upset)	0.59
Self-image (SI), 4 items	SI3 (good teeth)	0.69
	SI4 (good about myself)	0.55
	SI5 (will have good teeth)	0.78
	SI6 (will have good health)	0.39

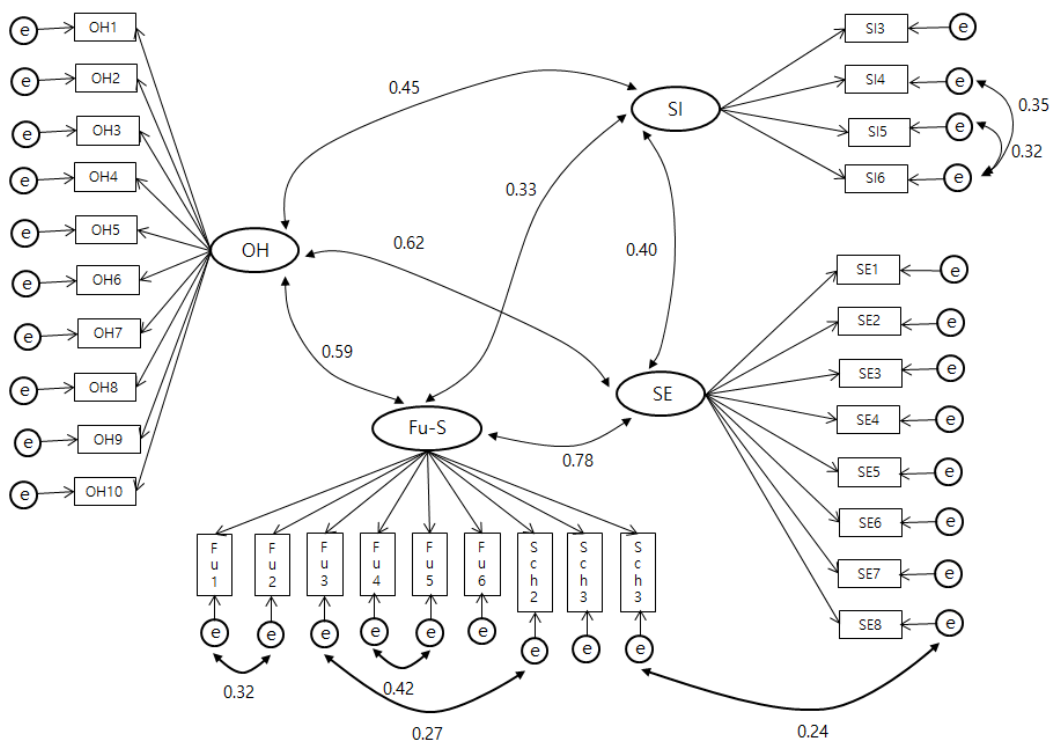


그림 18. 구강건강영향 프로파일(COHIP)의 공분산 기반 2차 계층모형(Cho et al, 2016)

약어설명: OH: Oral Health, Fu-S: Functional Well-being including school activities, SE: Social/Emotional Well-being, and SI: Self-image.

본 연구에서는 구강건강(Oral Health) 차원을 형성적 측정모형으로 가정하

고 4개의 1차 구성개념인 차원과 2차 구성개념의 관계가 형성적인 것으로 가정하였다. 그 이유는 하위 개념들이 서로 유사하거나 교환이 가능하지 않고, 각각이 구별되는 독특한 측면을 가지고 있으며, 형성적인 인과적 관계에 가깝기 때문이다. 본 연구에서는 구강건강을 제외한 1차 구성개념은 반영적 측정구조를 가지며 2차 구성개념은 형성적 특성을 가지는 2차 계층 구조방정식을 분석모형으로 선정하였다(그림 19).

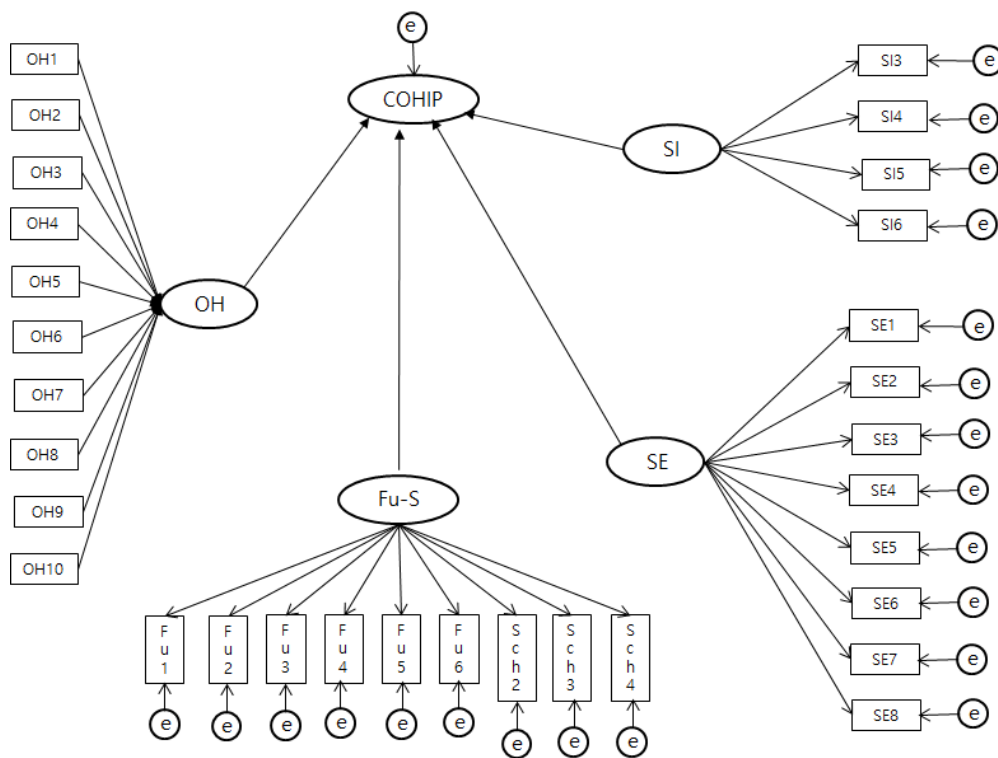


그림 19. 본 연구의 구강건강영향 프로파일(COHIP) 분석 모형

약어설명: COHIP: COHIP score, OH: Oral Health, Fu-S: Functional Well-being including school activities, SE: Social/Emotional Well-being, and SI: Self-image.

### 1.3.1 COHIP의 점수 계산방법

현재 각 요인 내에 속하는 문항의 점수를 더하여 차원 점수로 사용하고 전체 항목의 측정값을 더하여 COHIP 점수로 사용하는 것이 일반적이다(Broder et al. 2007a; Ahn et al, 2012). 이러한 합산점수를 사용하는 방법은 차원 내에서는 가중치를 각 문항 당 1로 부여하고 차원 간의 합산에 사용하는 가중치는 대략 차원 당 문항의 숫자와 유사한 값을 부여하는 것으로 해석할 수 있다.

## 1.4 어린이 일상활동 구강영향(Child Oral Impact of Daily Performances, COIDP)

### 1.4.1 COIDP 개요 및 구조

COIDP 도구는 먹기, 말하기, 치아 닦기, 잠자기, 정서적 안정, 웃기, 공부하기 및 사회적 활동 등 8개 항목으로 구성된 설문 도구로서, 2004년 Gherunpong 등에 의하여 영어로 개발된 후 스페인어, 불어, 이탈리아어, 포르투갈어, 및 탄자니아어 등으로 번역되어 타당성 검증을 거쳐 활용되고 있다. 표 9에 나타난 바와 같이 기본적으로 COIDP의 각 문항은 구강건강의 다양한 기능이나 정서 및 활동에 대한 영향 등의 상호 구별되는 측면을 나타내도록 구성되어 있으므로 COIDP는 형성적인 측정구조를 가지는 것으로 간주하였다.

표 10. 어린이 일상활동 구강영향(COIDP) 설문

항목	빈도	심도
씹기	지난 3개월 동안	지난 3개월 동안 구강상태로 인한 불편의 강도는?
말하기	구강상태로 인한 불편을	
치아 닦기	얼마나 자주	0.영향 없음. 1 (거의 영향 없었음) ~ 5 (매우 심각한 영향을 미쳤음)
잠자기	경험하였는지?	
감정 유지	1. 한 달 1회 미만 2.	
치아 보이며 웃기	한 달 1~2회 3. 일주일	
공부하기	1~2회 4. 일주일	
친구 만나기	3~4회 5. 거의 매일	

### 1.4.2. COIDP 점수 체계

COIDP는 지난 3개월 간 구강건강 관련하여 일상생활에 대한 불편이나 제한을 경험한 유무와 빈도 및 심도(severity)로 조사한다. 빈도는 1=한 달 1회 미만, 2=한 달 1~2회, 3=일주일 1~2회, 4=일주일 3~4회 및 5=거의 매일 등의 답지 중에서 선택하도록 하였고 심도는 '영향 없음'을 0점으로 부여하고, 거의 영향 없었음(1점)에서 매우 심각한 영향을 미쳤음(5점)의 범위에서 학생 본인이 점수를 선택하도록 하였다. 조사된 COIDP 설문결과는 (1) COIDP 영향 여부(indicator)와 (2) 각 항목의 빈도와 심도를 곱하여 구한 각 항목점수(item score) 및 (3) 각 항목점수의 합을 구한 후 이를 최고점수[8(항목)\*5(빈도)\*5(심도)=200]로 나누어 100을 곱하여 퍼센트로 계산한 COIDP 점수(COIDP score)로 분석에 이용한다.

#### 1.4.3 COIDP에 대한 기존의 분석 모형

대부분의 기존 연구에서는 심리측정적 방법론에 의거하여 요인분석을 실시하여 그 구조적 타당성을 검증하고자 노력하였다(Bernabe et al, 2008; Mtaya et al, 2007; 김혜영 등, 2010). COIDP 분석에는 심리측정적 방법론에 의거하여 각 문항의 내적일치도, 반영적 관계를 가정한 수렴타당도, 판별타당도 및 구조타당도 등이 연구된 바 있으나, 내적일치도가 낮고 요인구조의 분류가 적절하게 나타나지 않은 경우가 많았고 또한 여러 연구들 간의 차이가 커서 명료한 결론에 이르지 못하였다(Bernabe et al, 2008; Mtaya et al, 2007; Yusuf et al, 2006; 김혜영 등, 2010). 이는 근본적으로 도구의 특성에 맞지 않는 측정구조를 적용한 결과로 나타난 현상으로 사료된다.

#### 1.4.4 COIDP 도구 분석 모형

본 연구에서는 COIDP의 항목들이 각각 독특한 측면을 반영하고 있어 상호

교환되지 않고 항목이 원인이 되어 결과적으로 2차 구성개념에 영향을 주는 인과적 관계임을 고려하여 형성적 측정모형이 적합하다고 보아 그림 20과 같이 분석모형을 설정하였다.

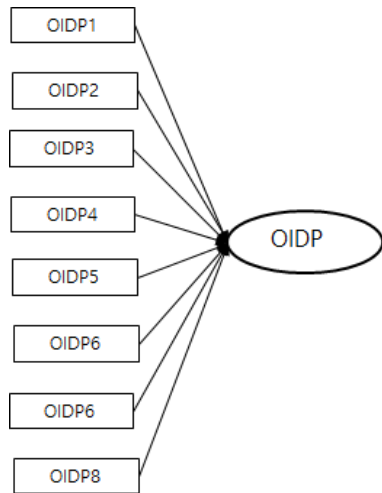


그림 20. 본 연구의 일상활동 구강건강영향(COIDP) 분석 모형

## 2. 반영적 측정모형 기반 데이터 생성 및 분석(모의실험)

모의실험은 반영적 관계와 형성적 관계를 포괄하여 용이하게 분석이 가능한 PLS-SEM을 적용하고자 할 때 원리적으로 상이한 반영적 측정모형의 구조를 가진 데이터에 대하여 나타날 수 있는 추정오류의 정도를 가늠하고자 설계되었다. 모의실험에서는 요인계수의 크기, 지표문항의 수, 표본 크기 등의 조건을 달리하여 반영적 측정구조를 가지는 데이터를 반복 생성하고 이를 CB-SEM과 PLS-SEM의 방법으로 각각 분석한 결과를 비교하여 PLS-SEM과 관련된 비일관성(inconsistency)을 평가하고자 하였다. 대상 자료가 1차 측정모형에 국한되므로 CB-SEM과 PLS-SEM의 구체적 적용으로 CFA와 PCA가 각각 분석에 사용되었다.

### 2.1 반영적 측정모형 기반 데이터 생성 및 분석

반영적 측정모형 기반 데이터의 생성은 다음과 같은 조건을 부여하여 수행되었다

- 1) 요인계수: low(.4), medium(.6), large(.8)
- 2) 차원 당 문항 수: 3, 6, 12 개
- 3) 표본 크기: 150, 300, 500, 1000개 표본
- 4) 반복횟수: 1000

위의 조건을 복합적으로 하여 총  $3*3*4=36$  조건을 부여하여 데이터 생성 및 분석을 실시하였다.

#### 2.1.1 데이터 생성 및 공통요인분석법 분석

반영적 데이터 생성 및 CFA를 이용한 반영적 측정모형 분석에는 MPlus 버전 7.3을 사용하여 Monte Carlo simulation 과정을 수행하였다. 오차항의 조건으로는 1-요인계수의 제곱을 적용하여 요인계수가 0.4 이면  $1-0.4^2=0.84$ , 0.6이면 0.64, 0.8이면 0.36 으로 각각 부여하였다. 요인점수는 평균을 0으로 분산은 1로 하여 표준화하였다.

MPlus 버전 7.3을 사용하여 CFA분석을 실시한 결과, 1000개의 주성분 요인계수의 평균값과 표준편차를 얻고 이를 이용하여 추정된 요인계수와 참값의 차이, 추정된 요인계수의 변동성을 나타내는 표준오차(Standard Error, SE) 및 평균 제곱오차(Mean Squared Error, MSE)를 산출하여 비교하였다(그림 21, 그림 22 및 그림 23).

다음은 분석을 위한 MPlus 프로그램 예시이다(6개 변수, 요인계수는 0,6, 표본크기는 150 인 경우).

```

MONTECARLO:
  NAMES ARE y1-y6 ;
  NOBSEVATIONS = 150 ;
  NREPS = 1000 ;
  SEED = 4533 ;
  REPSAVE = ALL ;
  SAVE = EX12_1*.DAT ;
  RESULTS = EX12_1.SAV ;

MODEL POPULATION:
  f BY y1-y6@.6 ;
  f@1 ;
  [f@0] ;
  y1-y6@.64;

MODEL: f BY y1-y6* ;
  f@1 ;
  [f@0] ;

OUTPUT: TECH9 ;

```

### 2.1.2 데이터 생성 및 주성분분석법 분석

데이터 생성 및 PCA를 이용한 반영적 측정모형 분석을 위하여 먼저 MPlus 버전 7.3을 사용하여 데이터를 생성하고 저장하여 Python 프로그램 이용하여 1000개의 데이터 파일을 하나로 합친 후(by rep) SAS 통계프로그램의 PROC FACTOR를 이용하여 PCA를 시행하였다. 그 결과 1000개의 주성분 요인 계수의 평균값과 표준편차를 얻고 이를 이용하여 추정된 요인계수와 참값의 차이, 추정된 요인계수의 SE 및 MSE를 산출하여 비교하였다(그림 21, 그림 22 및 그림 23).

다음은 분석을 위한 SAS 프로그램이다.

```
filename mcfile1 'H:\Documents\dat.csv';

data mc3_1_1;
infile mcfile1 dlm=';';
INPUT rep y1- y12 ;
RUN;

ods trace on / listing;

PROC FACTOR DATA=mc3_1_1 method=prin nfact=1 noprint outstat=pattern ;
var y1 - y12 ;
by rep;
run;

ods trace off;

proc sort data=pattern; by rep;
data loadings;set pattern;
by rep;
if last.rep;
RUN;

proc univariate data=loadings;
run;
```

### 2.2 분석 결과의 평가

반영적 측정구조를 기반으로 1000회 반복하여 생성된 데이터에 대하여

CB-SEM원리에 맞는 CFA를 실시하고 또한 PLS-SEM의 원리에 맞는 PCA를 실시하여 각각의 방법을 통하여 산출된 요인계수 추정치와 데이터 생성 시에 정한 요인계수를 비교하였다.

일반적으로 점 추정치를 평가함에 있어서 편향(bias), SE 및 MSE를 사용하여 평가하게 된다(Ross, 2012). 만약 모수  $\theta$ 를 가지고 있는 분포에서 표본으로  $\mathbf{X}=(X_1, \dots, X_n)$ 를 뽑고, 모수  $\theta$ 를  $d=d(\mathbf{X})$ 를 이용하여 추정하였다고 하자. 이때 편향과 추정 요인계수 SE 및 MSE의 관계는 다음과 같이 나타낼 수 있다.

$$\begin{aligned}
 MSE &= E[(d(X) - \theta)^2] && \text{(식 4.1)} \\
 &= E[(d - E[d] + E[d] - \theta)^2] \\
 &= E[(d - E[d])^2] + (E[d] - \theta)^2 \\
 &= \text{variance}(d) + \text{bias}^2
 \end{aligned}$$

식 4.1에 의하면 MSE는 SE의 제곱과 편향의 제곱의 합으로 나타난다. 이 연구에서는 모수의 참값을 고정하여 다수의 데이터를 생성하고 일정한 방법으로 이를 분석하여 모수의 점추정치를 얻었고 추정치가 모수를 잘 추정하는지 평가하기 위하여 추정요인계수의 참값과의 차이인 편향(bias), 추정 요인계수 SE 및 MSE를 비교의 대상으로 적용하였다.

#### 1) 추정요인계수와 요인계수 참값과의 차이

반영적 측정구조의 특성을 가지는 자료에 대하여 그 원리를 구현하는 CFA를 적용한 결과 모든 경우의 조건에서 PCA를 적용한 경우에 비하여 요인계수의 참값과 추정요인계수의 평균과의 차이가 더 작았다. CFA를 적용하였을 때 모든 경우의 조건에서 0.04미만의 차이를 보였다

반면 반영적 측정구조의 특성을 가지는 자료에 대하여 PCA를 적용한 경우 요인계수의 참값과 추정요인계수의 평균과의 차이가 0.02~0.26의 범위로 큰 차이

를 나타내었다. 이는 대상 자료의 생성 원리와 통계적 추정 원리가 상이함에 따라 모수의 편향(bias)이 나타난 것으로 볼 수 있으며, 요인계수가 작을수록, 차원 당 문항 수가 작을수록, 표본 크기가 작을수록 큰 편향을 보이는 경향이 있다. 편향의 크기 0.1을 기준으로 보면, 차원 당 문항 수가 3개 일 때에는 요인계수가 0.6이하인 경우에만 표본 크기와 무관하게 편향의 크기가 0.1 보다 작았다. 차원 당 문항 수가 6개 일 때에는 표본 크기와 무관하게 요인계수가 0.6이상인 경우에 편향의 크기가 0.1 보다 작았고 차원 당 문항 수가 12개 일 때에는 모든 경우에 편향의 크기가 0.1 보다 작았다(그림 21).

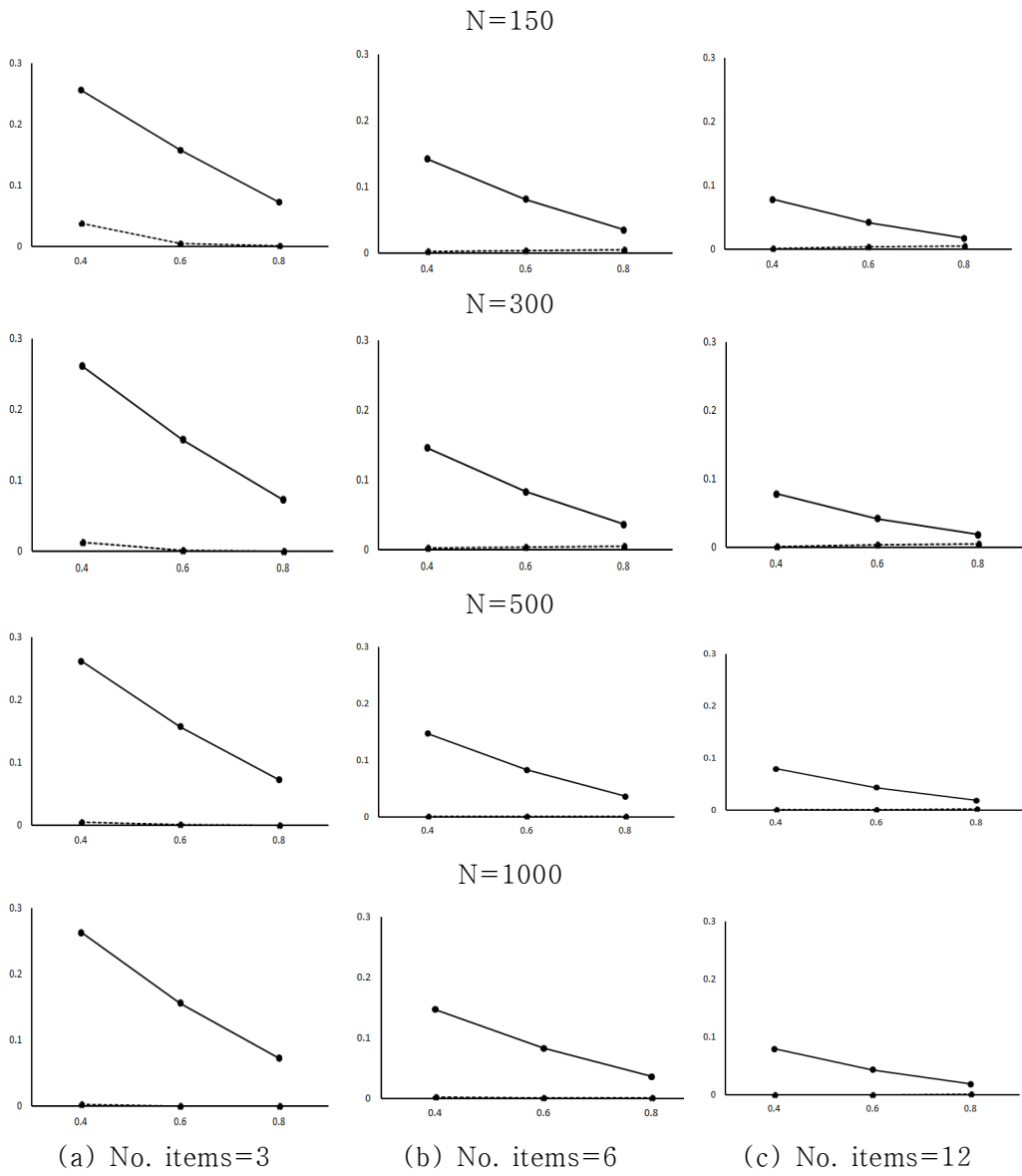


그림 21. 요인계수 참값과 추정치 평균의 차이(요인분석법: 점선, 주성분분석법: 실선; 가로축: 요인계수 참값, 세로축: 평균 요인계수의 차이)

## 2) 추정 요인계수의 표준오차

추정 요인계수의 표준오차는 모든 경우의 조건에서 CFA를 적용한 경우 PCA를 적용한 경우에 비하여 추정 요인계수의 표준오차가 더 큰 것으로 나타나 PCA가 CFA 보다 더 안정적인 추정 결과를 나타내었다. 전반적으로 표본의 크기가 커지면 표준오차의 크기가 급격하게 감소하였다. 또한 요인계수가 클수록, 차원 당 문항 수가 많을수록 표준오차의 크기가 감소하였다.

PCA 적용 시에는 거의 모든 경우 0.1 미만의 추정 요인계수의 표준오차를 보였으나 CFA를 적용하였을 때에는 차원 당 문항 수가 3개 이고 표본 크기가 300이하이면서 요인계수 참값이 0.4인 경우 요인계수의 표준오차가 0.1을 초과하였다. 그 이외의 모든 경우에는 요인계수의 표준오차가 0.1 보다 작아서 안정적으로 추정이 시행된 것으로 나타났다(그림 22).

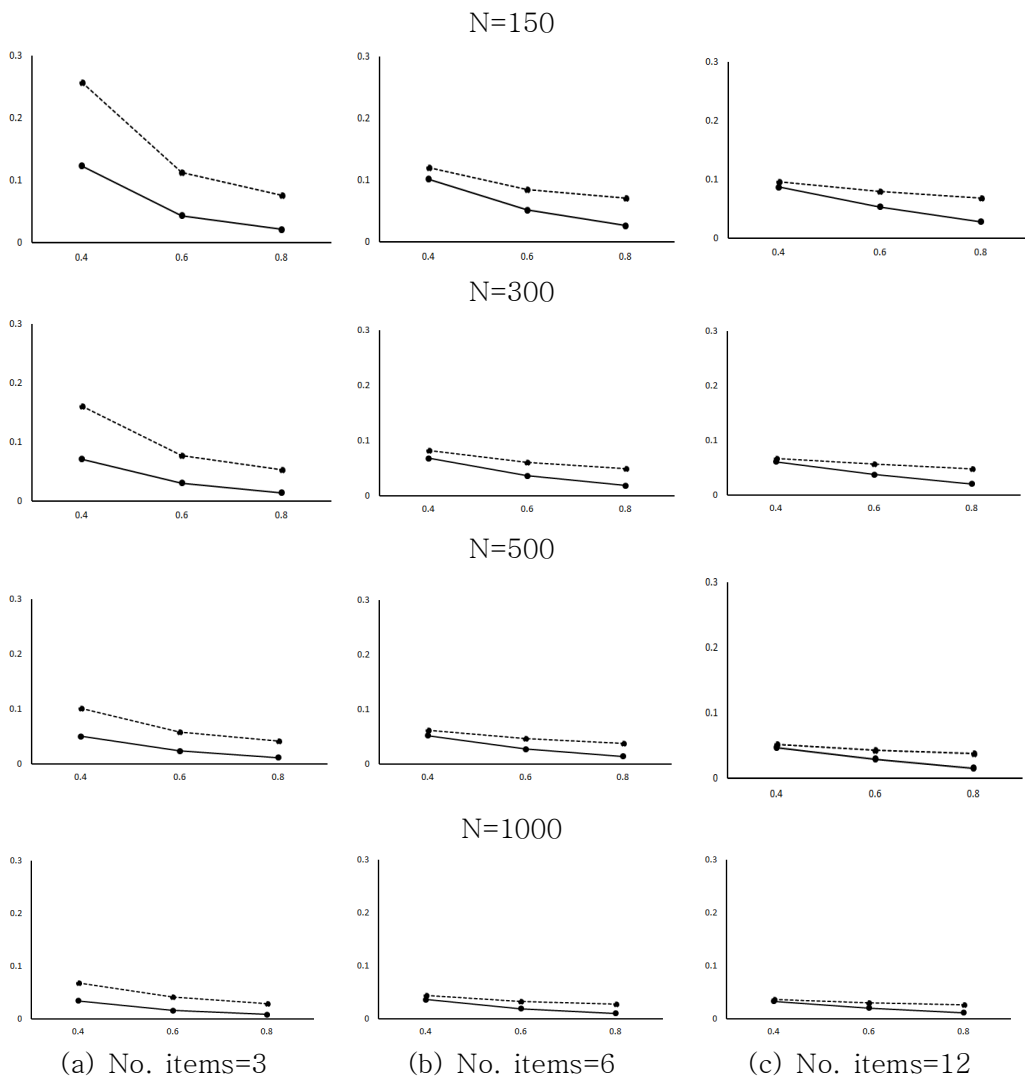


그림 22. 요인계수 추정치의 표준오차(요인분석법: 점선, 주성분분석법: 실선; 가로축: 요인계수 참값, 세로축: 평균 표준오차)

### 3) 추정 요인계수의 평균제곱오차(Mean Square Error, MSE)

추정 요인계수의 MSE는 대체로 CFA를 적용한 경우 PCA를 적용한 경우에 비하여 더 작았다. 요인계수가 클수록, 차원 당 문항 수가 많을수록 표본크기가 클수록 MSE의 크기가 감소하였으나 표본크기에 따른 MSE 변화는 미미한 편이었다. MSE 0.05를 상회한 경우는 차원 당 문항 수가 3개이고 요인계수가 0.4인 상태에서 주성분분석을 실시한 경우와 요인계수 0.4,이고 표본크기 150인 경우 요인분석을 실시한 경우에 한하여 나타났고 그 이외의 모든 경우에는 평균제곱오차 0.05 미만으로 나타났다(그림 23).

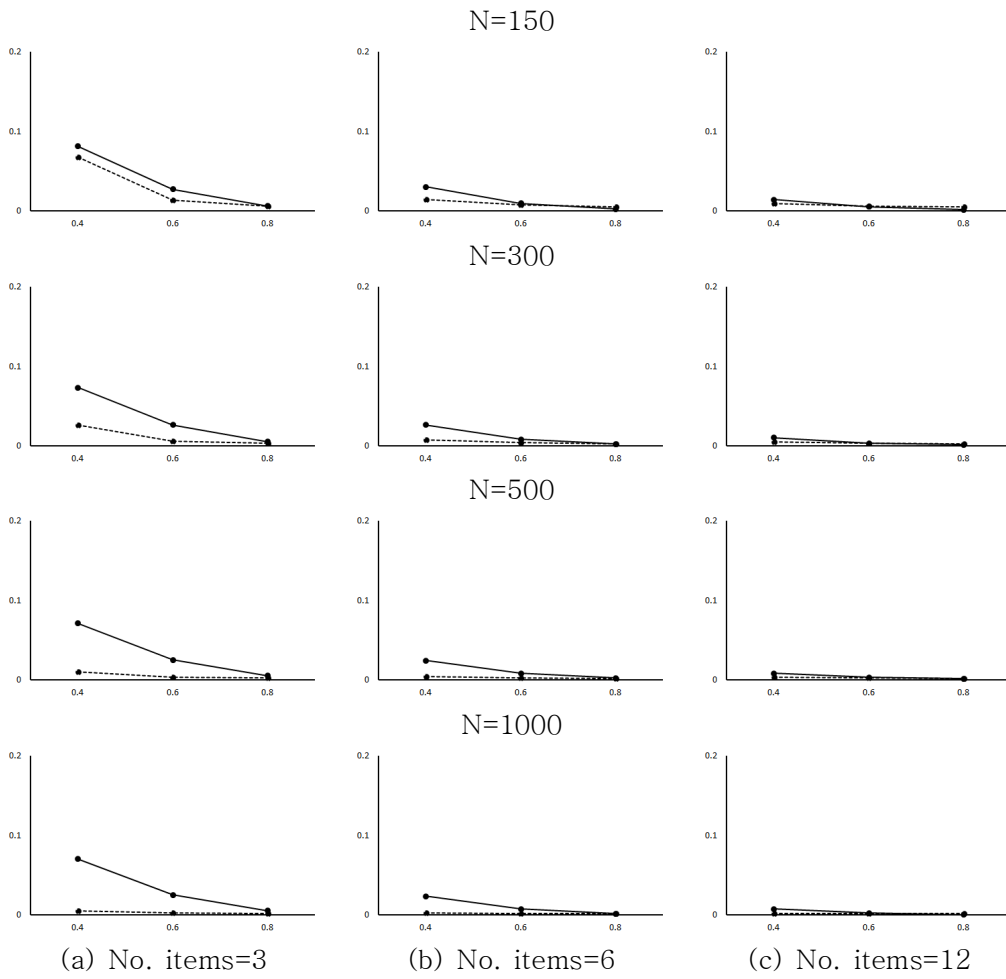


그림 23. 요인계수 추정치의 평균제곱오차(MSE) (요인분석법: 점선, 주성분분석법: 실선; 가로축: 요인계수 참값, 세로축: 평균 MSE )

주성분 분석의 대략적인 결과는 표 11로 정리되었다.

표 11. 반영적 자료의 주성분 분석 결과 평가

요인계수	영향 요인	요인계수	참값과의 차이	요인계수	표준오차
0.4	지표 변수 수	표본 크기에 무관하게 3개, 6개, 12개를 가진 경우	평균 0.26, 약 0.14, 약 0.07로 큰 편향이 나타남.	지표 변수 수에 따른 차이가 거의 없고	대략 0.03~0.1 범위를 보임.
	표본 크기	표본 크기에 거의 영향을 받지 않음		모든 조건에서	표본이 커지면 감소
0.6	지표 변수 수	표본 크기에 무관하게 3개, 6개, 12개를 가진 경우	평균 0.16, 약 0.08, 0.04로 나타남.	지표 변수 수에 따른 차이가 거의 없고	대략 0.02~0.05 범위를 보임.
	표본 크기	표본 크기에 거의 영향을 받지 않음		모든 조건에서	표본이 커지면 감소
0.8	지표 변수 수	표본 크기에 무관하게 3개, 6개, 12개를 가진 경우	평균 0.07, 0.04, 0.02로 나타남.	지표 변수 수에 따른 차이가 거의 없고	대략 0.01~0.03 범위를 보임.
	표본 크기	표본 크기에 거의 영향을 받지 않음		모든 조건에서	표본이 커지면 감소

#### 4) 반영적 데이터의 주성분분석(PCA) 시 오류 정도

반영적 측정모형에 기반 하여 데이터를 생성한 후 이를 PCA분석을 실시한 경우, 일정한 조건에서 추정된 요인계수의 편향이 크게 나타났으나 요인계수 추정치의 표준오차는 요인분석 시에 비하여 대부분 작아 안정적인 값을 보였다.

추정된 요인계수의 편향은 요인 당 지표변수가 3개인 경우 대부분 평균

0.1을 초과하여 컸고, 단 요인계수 참값이 0.8 이상인 경우에는 0.1 미만이었다. 또한 요인 당 지표변수가 6개이면서 요인계수 참값이 0.4인 경우에도 추정된 요인계수의 편향은 0.1을 초과하나 요인계수 참값이 0.6인 경우에는 편향이 비교적 작았다. 따라서 PCA분석 결과 요인계수 추정치의 편향을 작게 하려면 요인 당 지표변수의 수가 3개 보다 큰 조건이 필요하며 또한 요인계수 참값이 0.4를 초과하는 조건이 요구된다.

표본의 크기는 특히 요인 당 지표변수가 작고 요인계수 참값이 작은 경우에 추정치의 표준오차의 감소에 기여하였다.

## 2.3 SF-36과 COHIP 경험적 자료와의 비교

반영적 측정모형에 기반 하여 생성된 가상 자료에 대한 공통분산 및 주성분 분석 결과, 두 분석방법에 따른 차이가 특히 요인 당 지표변수 수와 요인계수의 크기에 따라 다르게 나타남을 확인한 바 있다. 이후 반영적 및 형성적 특성을 모두 가진 여러 건강관련 삶의 질 자료를 PLS-SEM방법으로 분석할 예정이므로 분석 대상이 될 반영적 측정모형 기반의 자료들이 어떤 조건을 가지고 있는지를 확인하고 그 결과의 정확도를 가늠해 볼 필요가 있다. 따라서 분석대상 자료 중 반영적 측정모형을 가진 SF-36과 COHIP 자료의 요인계수의 크기, 지표변수 수 및 표본 크기를 고려하여 이후 PLS-SEM을 적용할 때 나타날 요인계수 추정치의 편향과 추정치의 표준오차 및 MSE의 크기를 예측하였다.

### 2.3.1 SF-36 예시 자료의 조건과 예상 정확도

SF-36 예시 자료는 성인 남성 188명에 대한 자료로서 평균 요인계수는 0.72이며 평균 지표변수 수 4.38로서 비교적 편향이 작고 안정적인 추정을 수행할 것으로 예상된다(표 12). 앞의 모의실험 결과를 표시한 그림 21(110쪽), 그림 22(112쪽), 그림 24(114쪽) 및 표 11(115쪽)에 나타난 결과를 참고하여 해당 조건의 편향 정도를 가늠하면, 예상 평균 요인계수 편향은 대략 0.07~0.08의 범위를 예상할 수 있고, 추정 요인계수의 표준 오차는 대략 0.02~0.03 범위로, MSE는 0.01 이하가 될 것으로 생각된다.

표 12. SF-36 예시 자료의 요인계수의 크기 및 지표변수 수

반영적 요인	평균 요인계수*	지표변수 수
PF	0.58	10
RP	0.77	4
BP	0.81	2
GH	0.66	5
VT	0.67	4
SF	0.75	2

RE	0.85	3
MH	0.68	5
Mean	0.72	4.38

\*: 공분산 기반 구조방정식에 의한 표준화 요인계수임.

### 2.3.2 COHIP 예시 자료의 조건과 예상 정확도

COHIP 예시 자료는 어린이 2,236명에 대한 자료로서 평균 요인계수는 0.61이며 평균 지표변수 수 7.00으로서 비교적 편향이 작고 안정적인 추정이 가능할 것으로 예상된다(표 13). 앞의 모의실험 결과를 표시한 그림 21(110쪽), 그림 22(112쪽), 그림 24(114쪽) 및 표 11(115쪽)에 나타난 결과를 참고하여 해당 조건의 편향 정도를 대략 가늠하면, 예상 평균 요인계수 편향은 대략 0.08 정도를 예상할 수 있고, 추정 요인계수의 SE는 대략 0.02 정도로, MSE는 0.01 이하가 될 것으로 기대된다.

표 13. COHIP 예시 자료의 요인계수의 크기 및 지표변수 수

반영적 요인	평균 요인계수*	지표변수 수
Functional-School	0.56	9
Social-Emotional	0.68	8
Self Image	0.60	4
Mean	0.61	7.00

\*: 공분산 기반 구조방정식에 의한 표준화 요인계수임.

### 3. 건강 관련 삶의 질 측정 자료에 대한 PLS-SEM 분석

HRQoL 자료의 실증적 분석에는 SF-36, EQ-5D, COHIP 및 COIDP가 사용되었으며 PLS-SEM을 구현하는 통계 소프트웨어는 SmartPLS(Hair et al, 2014)를 사용하였다.

#### 3.1 반영적-형성적 측정구조: SF-36

##### 3.1.1 SF-36 조사 자료

실증 분석에 사용된 SF-36 자료는 이차 자료로서 차보경 등(2016)의 연구에서 사용된 자료를 저자의 협조를 얻어서 사용하게 되었다. 연구대상자는 성인 남성 196명의 자료 중 결측치가 있는 사례를 삭제하여 결측치가 없는 188건의 자료가 최종적으로 모형구축에 사용되었다.

##### 3.1.2 SF-36 분석 모형

PLS-SEM 모형을 이용한 SF-36 분석은 그림 9에 나타난 이론적 모형에 따라 35개 지표변수와 8개의 1차 구성개념 및 1개의 2차 구성개념을 설정하고 반복지표변수 모형을 적용하고 반영적 Mode A를 이용하여 수행하였다(그림 24). 1차 구성개념이 반영적이고 2차 구성개념이 형성적 관계로 설정된 상태에서 형성적 Mode B를 적용한 결과 부트스트랩 과정이 수행되지 않았고 이에 따라 반영적 Mode A를 적용하였다.

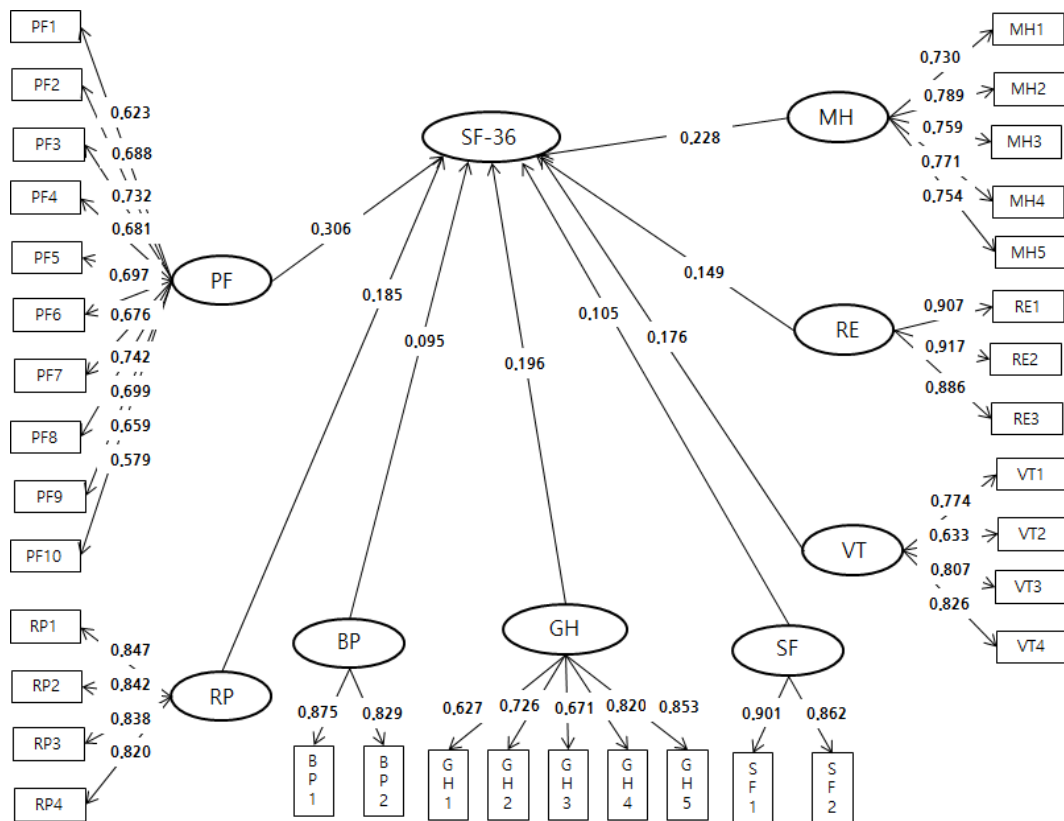


그림 24. SF-36 측정도구의 PLS-SEM 분석 모형과 분석 결과

약어설명 - SF-36: SF-36 score, PF: Physical Functioning, RP: Role limitation due to physical problems, BP: Bodily pain, GH: General health perceptions, SF: Social functioning, VT: Vitality, RE: Role limitation due to emotional problems, and MH: Mental health.

표 14에 나타난 바와 같이 1차 구성개념과 2차 구성개념 간의 구조계수는 모두 유의하게 나타났고, 2차 구조식이 형성적 구조를 가져서 1차 구성개념이 2차 구성개념을 모두 설명하므로 결정계수는 1이 나왔다(Becker et al, 2012).

표 14. SF-36 에 대한 반복지표변수 모형의 구조계수(n=188)

	Structural Coefficients	
	Estimate	P-value
Physical Functioning	0.306	<.001
Role Physical	0.185	<.001
Bodily Pain	0.095	<.001
General Health	0.196	<.001
Vitality	0.176	<.001
Social Functioning	0.105	<.001
Role Emotional	0.149	<.001
Mental Health	0.228	<.001

SF-36의 R-square=1.00

SF-36의 반영적 측정모형으로 이루어진 1차 구조를 평가해 보면 먼저 내적 일관성 면에서 복합신뢰도(Composite Reliability, CR)가 대략 0.6~0.9 사이 이면 수용 가능한 것으로 볼 수 있어서 대부분의 1차 구성개념이 그 범위에 들어가나 단지 정서적 이유로 인한 역할제한(Role Emotional)의 경우 0.93으로 큰 값으로 나와서 의미상으로 중복되는 불필요한 항목이 사용되었는지 살펴볼 필요성이 있다(Nunally & Bernstein, 1994). 역할제한은 정서적인 문제로 일하는 시간이나 일의 양을 줄이거나 일하는데 평소처럼 주의를 기울이지 못하였다 등 세 문항으로 이루어진 차원으로 정서적인 문제가 일하는 시간, 수행능력 및 주의 집중 등에 비슷한 정도로 큰 영향을 주는 것으로 해석할 수 있다 .

지표변수 신뢰성 면에서는 요인계수 0.7 이상을 기준으로 할 때 신체적 기능(Physical Functioning)과 일반건강(General Health) 구성개념의 경우 요인계수가 0.7 미만인 지표변수들이 존재하였다. 평균추출분산(average variance extracted, AVE)을 살펴보면 0.5 이상의 기준을 대부분 충족하나 신체적 기능의 경우만 0.46으로 기준에 미달하였다(표 15).

PLS-SEM과 CB-SEM에 의한 반영적 측정모형의 추정에 있어서 요인계수의 추정치 차이는 0.09(= 0.77-0.68)로 나타내데 실제 요인계수의 추정치 차이는

상당하여 35 개 지표변수 중 20 개가 0.1 에 가깝거나 0.1 이상의 차이를 보였다 (표 14). 특히 AVE가 0.46으로 낮고 평균 요인계수가 0.6에 미달하는 신체기능 차원에서 큰 차이를 보였는데 이는 차원을 공통적으로 반영하는 공통분이 적고 상대적으로 오차분산이 큰 경우이므로 오차분산을 분리하지 않고 추정과정에 전체 분산을 반영하는 PLS-SEM의 특성에 기인한 것으로 볼 수 있다.

표 15. SF-36의 PLS-SEM 결과 평가 및 CB-SEM 결과와의 비교(n=188)

Domain	Indicators	PLS-SEM*			CB-SEM#
		요인 계수	AVE	복합신 뢰도	요인계수
Physical functioni ng	PF1(Vigorous activities)	0.62	0.46	0.90	<b><u>0.25</u></b>
	PF2(Moderate activities)	0.69			<b><u>0.41</u></b>
	PF3(Lifting/carrying groceries)	0.73			<b><u>0.48</u></b>
	PF4(Climbing several steps of stairs)	0.68			<b><u>0.36</u></b>
	PF5(Climbing one step of stairs)	0.70			0.71
	PF6(Bending, kneeling or stooping)	0.68			<b><u>0.44</u></b>
	PF7(Walking more than a mile)	0.74			<b><u>0.59</u></b>
	PF8(Walking several blocks)	0.70			<b><u>0.92</u></b>
	PF9(Walking one block)	0.66			<b><u>0.91</u></b>
	PF10(Bathing or dressing yourself)	0.58			<b><u>0.75</u></b>
	Mean(PF)	0.68			<b><u>0.58</u></b>
Role Physical	RP1(Cut down amount of time)	0.85	0.70	0.90	0.88
	RP2(Accomplished less )	0.84			0.86
	RP3(Limited in the kind of work )	0.84			<b><u>0.69</u></b>
	RP4(Had difficulty in working)	0.82			<b><u>0.65</u></b>
	Mean(RP)	0.84			<b><u>0.77</u></b>
Bodily Pain	BP1(How much bodily pain)	0.88	0.81	0.89	<b><u>0.69</u></b>
	BP2(Pain interfere with work)	0.93			0.92
	Mean(BP)	0.91			0.81
General Health	GH1(In general your health is)	0.67	0.55	0.86	0.59
	GH2(Get sick easier)	0.63			<b><u>0.49</u></b>
	GH3(Healthy as anybody)	0.82			<b><u>0.72</u></b>
	GH4(Expect health to get worse)	0.72			0.65
	GH5(My health is excellent)	0.85			0.86
	Mean(GH)	0.74			0.66
Vitality	VT1(Feel full of pep)	0.81	0.58	0.85	<b><u>0.68</u></b>
	VT2(Have a lot of energy)	0.83			0.78

	VT3(Feel worn out)	0.77			0.70
	VT4(Feel tired)	0.63			<b>0.51</b>
	Mean(VT)	0.76			<b>0.67</b>
Social	SF1(Extent-interfere social )	0.86	0.78	0.88	<b>0.69</b>
Function	SF2(Time-interfere social )	0.90			0.81
	Mean(SF)	0.88			0.75
Role	RE1((Cut down amount of time)	0.91	0.82	0.93	0.85
Emotio	RE2(Accomplished less )	0.92			0.89
nal	RE3(Not careful )	0.89			0.81
	Mean(RE)	0.91			0.85
	MH1(Been nervous)	0.73	0.57	0.87	<b>0.62</b>
Mental	MH2(Felt down)	0.79			<b>0.70</b>
Health	MH3(Felt calm and peaceful)	0.77			0.73
	MH4(Felt downhearted)	0.76			<b>0.65</b>
	MH5(Been happy)	0.75			0.73
	Mean(MH)	0.76			0.69
평균		0.77			0.68

\*: 본 연구의 PLS-SEM 적용 결과, #: CB-SEM 분석 결과

Fornell-Larker 기준에 따라 판별타당도를 검증하기 위하여 표 16이 작성되었다(61쪽 판별타당도 설명 참조). 대각선에는 각 구성개념의 평균추출분산(AVE) 제곱근을 배치하고 대각 외에는 다른 구성개념과의 상관계수를 배치하여 상호 비교한 결과 모든 경우에 대각선의 AVE 값이 대각 외의 상관계수에 비하여 커서 판별타당도를 만족하였다(Fornell and Larker, 1981). 또한 지표변수가 속한 구성개념에 대한 요인계수가 다른 구성개념에 대한 cross-loading에 비하여 모든 경우에 커서 역시 판별타당도 기준을 충족하였다.

표 16. SF-36모형의 판별타당도(Fornell-Larker 기준)

	PF	RP	BP	GH	VT	SF	RE	MH
PF	0.67							
RP	0.42	0.84						
BP	0.41	0.36	0.90					
GH	0.40	0.26	0.34	0.74				
VT	0.30	0.26	0.40	0.68	0.76			
SF	0.25	0.48	0.38	0.32	0.50	0.88		
RE	0.22	0.56	0.23	0.20	0.32	0.57	0.90	
MH	0.29	0.40	0.40	0.55	0.73	0.58	0.47	0.76

PF: Physical Functioning, RP: Role limitation due to physical problems, BP: Bodily pain, GH: General health perceptions, SF: Social functioning, VT: Vitality, RE: Role limitation due to emotional problems, and MH: Mental health.

SF-36 도구의 이차구조의 결정계수는 형성적 이차 관계의 특성 상 1.0으로 나타났고 전체 구조모형의 적합도를 평가하는 global goodness of fit (GoF) 은 0.813 으로 큰 GoF의 기준인 0.36을 크게 상회하므로 적합도가 매우 좋다고 평가할 수 있다.

$$GoF = \sqrt{AVER^2} = \sqrt{0.661*1} = 0.813$$

표 17에는 SF-36 2차 계층모형의 PLS-SEM 추정과 관련한 평가결과를 요약하였다.

표 17. SF-36 2차 계층모형의 PLS-SEM 추정 결과 평가

평가 단계	평가 사항	평가 기준	SF-35 모형 평가
반영 적 측정 모형	내적 일관성 / 구성개념 신뢰성	CR 0.6~0.7 이상이면 수용가능. 0.9초과 시 주의	모든 구성개념에서 0.8 상회함
	지표변수 신뢰	외부 요인계수 0.7 이상	RE의 경우 0.93 대체로 만족. PF, GH의 일부

	성		지표변수가 기준에 미달함
평가	수렴타당도	AVE 0.5이상	대체로 충족됨. PF 0.46 로
	판별타당도	AVE값의 제곱근은 다른 구성개념 간의 상관관계 초과	미달
구조 모형 평가	결정계수 ( $R^2$ )	$R^2$ : small-0.02, medium-0.13, large-0.26	모두 충족됨
	Global goodness of fit	$GoF = \sqrt{AVE R^2}$ , small-0.1, medium-0.25, large-0.36	이차 형성적 구조이므로 1.0 0.81로 매우 만족.

CR: Composite Reliability, AVE: Average Variance Extracted, VIF: Variance Inflation Factor, GoF: Goodness of Fit, RE: Role limitation due to emotional problems, PF: Physical Functioning, and GH: General health perceptions.

### 3.2 형성적 측정구조: EQ-5D

#### 3.2.1 EQ-5D 분석 자료

국민건강영양조사 5기 1차년도인 2010년에 수집된 자료 총 8,388 중 필요한 항목에서 결측치를 보유한 사례를 제거하고 최종적으로 6,216건의 데이터를 분석에 적용하였다.

#### 3.2.1 EQ-5D PLS-SEM 모형

EQ-5D 삶의 질 도구의 형성적 측정구조를 적용하기 위하여 단일 변수로 전반적인 HRQoL을 EQ-VAS로 응답하게 한 EQ-VAS를 이용하여 모형화 하였다. 서로 다른 방법으로 특정된 유사한 두 구성개념 간의 관계를 규정한 상태로 PLS-SEM 모형을 이용하여 추정을 실시하였다(그림 25).

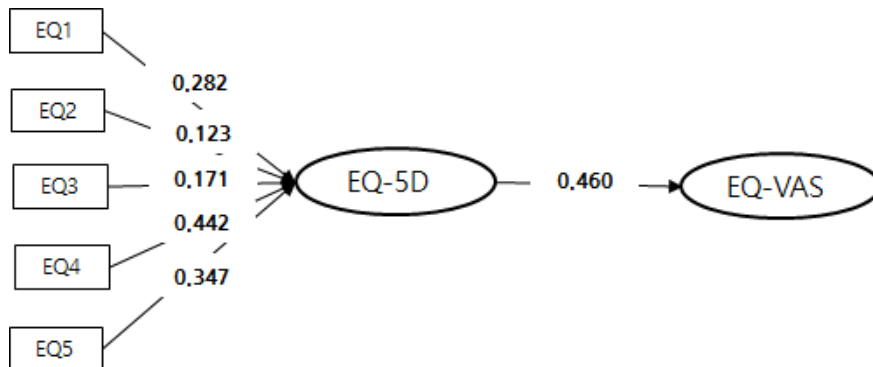


그림 25. EQ-5D 도구의 형성적 측정모형 PLS-SEM 분석

EQ-5D의 형성적 측정모형의 추정 결과 모든 외부 가중치가 유의한 것으

로 평가되었고 다중공선성의 문제가 발생하지 않는 것으로 나타났다(표 17). PLS-SEM과 PCA 방법에 의한 형성적 측정모형의 추정에 있어서 가중치의 추정치 차이가 상당한 정도로 나타나 다섯 개 지표변수 중 네 개가 0.1 이상의 차이를 보였다(표 18).

표 18. EQ-5D의 형성적 측정모형의 질 평가

Indicators	PLS-SEM			주성분 분석*
	가중치	P-value	VIF	가중치
Mobility	0.28	<.001	1.8	0.29
Self Care	0.12	0.001	1.4	<u>0.25</u>
Daily Life	0.17	<.001	2.0	<u>0.31</u>
Pain	0.44	<.001	1.6	<u>0.27</u>
Depression	0.35	<.001	1.1	<u>0.22</u>
Mean Weight	0.27			0.27

\*: 동일한 자료에 대하여 1 요인을 적용한 주성분 분석을 하여 나온 결과

EQ-5D 구성개념과 EQ-VAS 간의 구조계수는 0.46으로 유의한 값을 보였다(표 19). 만약 이 모형을 일종의 중복 분석으로 간주한다면 결정계수 0.212는 수렴타당도 면에서 만족하지 못하였다(표 20).

표 19. EQ-5D 모형과 EQ-VAS

	Structural Coefficients	
	Estimate	P-value
EQ-5D	0.460	<.001

EQ-VAS의 R-square=0.212    EQ-5D의 f-square=0.269

표 20. EQ-5D 형성적 측정모형 및 구조모형의 평가

	평가 항목	평가 기준	EQ-5D 모형 평가
형성적 측정모형	수렴타당도	중복 분석: 수정된	0.212 로서 기준보다 낮음.
		R <sup>2</sup> 값이 0.81, 0.64 초과	
	다중공선성	VIF가 5 초과 / 공차가 0.2 미만	모든 지표변수가 조건에 만족함
구조모형	외부 가중치 타당성	외부 가중치 유의	모두 유의함
	결정계수 (R <sup>2</sup> )	small-0.02, medium-0.13, large-0.26	0.212 로 큰 편임

VIF: Variance Inflation Factor

### 3.2.2 EQ-5D의 주성분 분석

위의 EQ-5D의 PLS-SEM 분석 결과와의 비교를 위하여 EQ-5D 자료에 대하여 주성분 분석을 실시하였다(그림 26). 주성분 분석 결과 나타난 우측의 가중치는 PLS-SEM에서 나타난 가중치와 다소 차이가 있음을 볼 수 있다. EQ-5D 모형에서 PLS-SEM과 주성분 분석의 차이는 EQ-VAS와의 설정 여부인데, PLS-SEM에서 지표변수 뿐 아니라 회귀식으로 연결된 변수를 고려한다는 점에서 차이가 나타난 것으로 보인다.

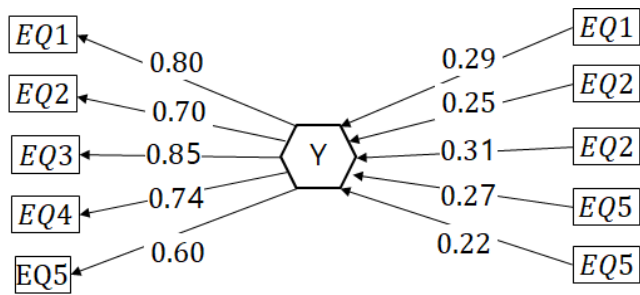


그림 26. EQ-5D 도구의 형성적 측정모형 주성분 분석

### 3.3 혼합적 측정구조: COHIP

#### 3.3.1 COHIP 분석 자료

Cho et al (2016)의 연구에 사용된 조사 자료를 분석에 이용하였다. 총 2,236건의 자료 중 결측치를 포함한 사례를 제거하고 난 후 2,068건의 자료가 최종적으로 분석에 이용되었다.

#### 3.3.2 PLS-SEM 모형을 이용한 COHIP 분석

PLS-SEM 모형을 이용한 COHIP 분석은 그림 27과 같이 이론적 모형에 따라 31개 지표변수와 1개의 1차 형성적 구성개념, 3개의 1차 반영적 구성개념 및 1개의 2차 구성개념을 설정하고 반복지표변수 모형 및 형성적 Mode B를 이용하여 수행하였다. 1차 구성개념이 형성적 및 반영적인 관계의 혼합이고 2차 구성개념이 형성적 관계로 설정된 상태에서 결과적으로 2차 구성개념의 복합점수를 추정과정으로 포함하는 형성적 Mode B를 적용함이 적절하다고 보았다.

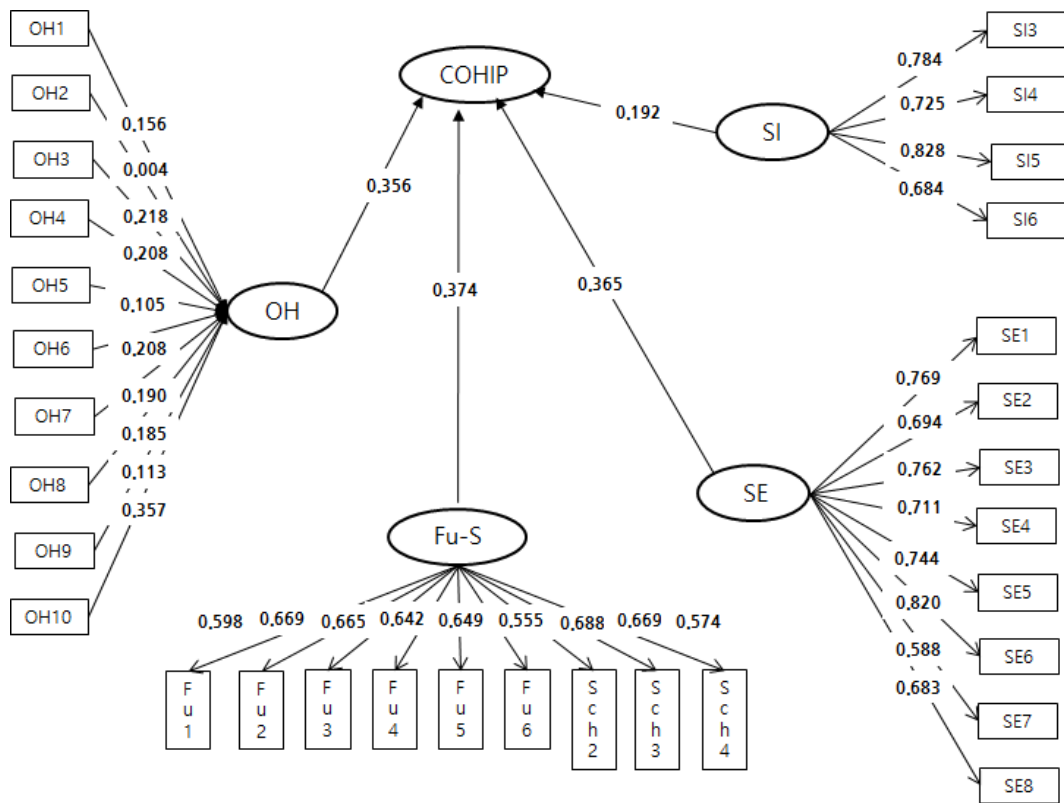


그림 27. COHIP 도구의 혼합적 측정모형 및 계층적 구조모형 분석

약어설명: OH: Oral Health, Fu-S: Functional Well-being including school activities, SE: Social/Emotional Well-being, and SI: Self-image.

표 21에 나타난 바와 같이 1차 구성개념과 2차 구성개념 간의 구조계수는 모두 유의하게 나타났고, 2차 구조식의 결정계수는 1이었다(Becker et al, 2012).

표 21. COHIP의 형성적 구조계수(반복지표변수 모형, Mode B)

	Structural Coefficients	
	Estimate	P-value
Oral Health	0.356	<.001
Function-School	0.374	<.001
Socio-Emotional	0.365	<.001
Self-Image	0.192	<.001

COHIP의 R-square=1.00

COHIP의 형성적 1차 구조인 ‘Oral Health’를 평가한 결과, 하나의 지표변수(OH2)의 외부 가중치가 0.05 유의수준에서 유의하지 않은 것으로 나타났고 나머지 아홉 개의 지표변수는 유의하였다. 다중공선성과 관련된 문제는 없었다(표 22). PLS-SEM과 PCA 방법에 의한 형성적 측정모형의 추정에 있어서 가중치의 추정치 차이는 크지 않은 것으로 나타나 열 개 지표변수 중 세 개가 0.1에 가깝거나 그 이상의 차이를 보였다(표 22).

표 22. COHIP의 형성적 1차 구조 ‘Oral Health’에 대한 질 평가

Dom ain	Indicators	PLS-SEM			주성분 분석
		가중치	P-value	VIF	가중치
Oral Heal th	OH1(pain)	0.36	<.001	1.3	<b><u>0.19</u></b>
	OH2(mouth breathing)	0.004	0.929	1.2	<b><u>0.18</u></b>
	OH3(discolored teeth)	0.22	<.001	1.2	0.16
	OH4(crooked teeth)	0.21	0.004	1.1	0.14
	OH5(sore spots)	0.11	<.001	1.2	0.18
	OH6(bad breath)	0.21	<.001	1.3	0.18
	OH7(bleeding gums)	0.19	<.001	1.2	0.17
	OH8(food sticking)	0.19	<.001	1.4	0.20
	OH9(cold sensitivity)	0.11	0.006	1.3	<b><u>0.20</u></b>
	OH10(dry mouth)	0.16	<.001	1.2	0.18
Mean Weight		0.18			0.18

\*: 동일 자료에 대하여 1 요인을 적용한 주성분 분석 결과

표 23은 COHIP의 1차 구성개념 중 반영적 측정모형으로 이루어진 3개 구성개념의 평가 사항을 정리한 것이다. 지표변수 신뢰성 면에서는 기능-학교 구성개념의 한 지표변수만이 요인계수가 0.65로 다소 작았고 다른 모든 요인계수는 0.7을 상회하였다. 평균추출분산(AVE)을 살펴보면 두 개의 차원은 0.5 이상의 기준을 충족하였으나 기능-학교 차원은 0.40으로 기준에 미달하였다. 내적 일관성 면에서 복합신뢰도(Composite Reliability, CR)는 0.84~0.90으로 모든 차원이 요구되는 조건을 충족하였다(Nunally & Bernstein, 1994). PLS-SEM과 CB-SEM 방법에 의한 반영적 측정모형의 추정에 있어서 요인계수의 추정치 차이는 상당하여 20 개 지표변수 중 8 개가 0.1 에 가깝거나 그 이상의 차이를 보였고 평균으로는 0.06의 차이를 보였다(표 23). 이러한 차이는 요인계수가 다른 지표변수에 비하여 작고 상대적으로 오차분산이 큰 경우에 주로 발생하는 경향을 보였는데 이는 오차분산을 제외하고 공통분산 만을 추정에 사용하는 CB-SEM과 오차분산을 포함한 지표변수의 전체 분산을 이용하여 추정하는 PLS-SEM의 추정법이 다름에 의하여 발생한 것이다.

표 23. COHIP의 반영적 1차 구조의 질 평가 기준

Domain	Indicators	PLS-SEM			CB-SEM*
		요인계수	AVE	복합신뢰도	요인계수
Functional - School	Fu1(trouble chewing)	0.60	0.40	0.86	0.57
	Fu2(difficulty eating)	0.67			0.61
	Fu3(trouble sleeping)	0.67			<b>0.57</b>
	Fu4(difficulty pronouncing)	0.64			<b>0.55</b>
	Fu5(difficulty being understood)	0.65			<b>0.56</b>
	Fu6(difficulty keeping clean)	0.56			0.49
	Sch2(difficulty paying attention)	0.67			0.67
	Sch3(not wanting to speak)	0.67			0.63
	Sch4(not wanting to go to school)	0.57			<b>0.42</b>
	Mean(Fu)	0.63			0.56

	SE1(unhappy)	0.65	0.53	0.84	0.73
Social-Emotional	SE2(felt worried)	0.69			0.68
	SE3(avoid smiling)	0.76			0.74
	SE4(felt different)	0.71			0.68
	SE5(worried)	0.74			0.71
	SE6(felt shy)	0.82			0.79
	SE7(teased)	0.59			0.52
	SE8(upset)	0.68			<b>0.59</b>
	Mean(SE)	0.71			0.68
Self Image	SI3(good teeth)	0.78	0.57	0.90	<b>0.69</b>
	SI4(good about myself)	0.73			<b>0.55</b>
	SI5(will have good teeth)	0.83			0.78
	SI6(will have good health)	0.68			<b>0.39</b>
Mean(SI)	0.76			0.60	
Mean	0.68			0.62	

\*: Covariance-SEM에 따른 결과(Cho et al, 2016)

Fornell-Larker 기준에 따라 COHIP 모형의 판별타당도를 검증하였다(표 24). 대각선에 위치한 평균추출분산(AVE) 제곱근에 비하여 대각 외에 위치한 다른 구성개념과의 상관계수를 비교한 결과 대각선의 AVE 값이 더 큰 사회-정서 장애(Social-Emotional)와 자신의 이미지(Self Image)는 판별타당도를 만족하였다(Hair et al, 2011). 그러나 기능-학교(Functional-School)는 사회-정서 장애(Social-Emotional)와 판별되지 않았다. 지표변수가 속한 구성개념에 대한 요인계수가 다른 구성개념에 대한 cross-loading에 비하여 모든 경우에서 크게 나타나 판별타당도를 충족하였다.

표 24. COHIP 모형의 판별타당도(Fornell-Larker 기준)

	Functional-School	Social-Emotional	Self Image
Fu-Sch	0.64		
SE	0.66	0.72	
SI	0.25	0.29	0.76

R-square: 1.0.

약어설명: Fu-S: Functional Well-being including school activities, SE:

Social/Emotional Well-being, and SI: Self-image.

COHIP 도구의 이차구조의 결정계수는 형성적 이차 관계의 특성 상 1.0으로 나타났고 전체 구조모형의 적합도를 평가하는 GoF는 0.833 으로 큰 GoF의 기준인 0.36을 크게 상회하므로 적합도가 매우 좋다고 평가할 수 있다.

$$\text{GoF} = \sqrt{\text{AVER}^2} = \sqrt{0.5*1} = 0.71$$

표 25에는 COHIP 2차 계층모형의 PLS-SEM 추정과 관련한 평가결과를 요약하였다.

표 25. COHIP 2차 계층모형의 PLS-SEM 추정과 관련한 평가

	평가 사항	평가 기준	COHIP 모형 평가
형 성 적	수렴타당도	중복 분석: 수정된 R <sup>2</sup> 값이 0.81, 0.64 초과	자료의 한계로 분석이 가능하지 않음
	다중공선성	VIF가 5 초과 / 공차가 0.2 미 만	만족
	외부 가중치의 유 의성	외부 가중치 유의성	하나 이외의 외부 가중치는 모두 유의.
반 영 적 추 정 모 형	내적 일관성 / 구 성개념 신뢰성	CR 0.6~0.7 이상이면 수용 가 능. 0.9 초과 시 주의	0.84~0.9 범위로 적절함
	지표변수 신뢰성	외부 요인계수 0.7 이상	한 차원의 외부 요인계수 기준 이하.
	수렴타당도	AVE 0.5이상	두 차원 만족. 하나는 기준 이하
구 조 모 형	판별타당도	AVE값의 제곱근은 다른 구성개 념 간의 상관관계 초과	두 차원 만족. 하나는 만족하지 못함.
	결정계수 (R <sup>2</sup> )	R <sup>2</sup> : small-0.02, medium- 0.13, large-0.26	이차 형성적 구조이므로 1.0
	Global goodness of fit	GoF = $\sqrt{AVE \cdot R^2}$ , small-0.1, medium-0.25, large-0.36	0.71로 매우 만족.

CR: Composite Reliability, AVE: Average Variance Extracted, VIF: Variance Inflation Factor, GoF: Goodness of Fit.

### 3.4 형성적 측정구조: COIDP

#### 3.4.1 COIDP 자료

김혜영 등(2010)의 연구에 사용된 자료로서 초등학교 고학년 어린이 306명

의 조사 자료를 사용하였다. 이 자료에 COIDP 및 COHIP 도구가 둘 다 조사되어 있어서 모형 상 독립적으로 PLS-SEM 추정이 불가능한 형성적 측정구조인 COIDP를 정의하는데 COHIP를 이용하였다.

### 3.4.2 COIDP 도구의 평가와 분석

COIDP는 보통 인덱스로 지칭하며 도구의 내용 상 형성적인 측정구조일 것으로 여겨지고 있다. 여기서는 형성적 측정구조와 반영적 측정구조의 두 경우의 전제에 대해서 모두 분석하여 비교하고자 한다. 먼저 지표변수들 간의 스피어만 상관계수를 구한 결과 지표변수들 간의 상관이 중간 또는 작은 정도로 나타났다 (표 26).

표 26. COIDP 지표변수 간의 비모수적 상관계수(n=306)

	EAT	SPEAK	CLEAN	REST	EMOTION	SMILE	STUDY
EAT	1						
SPEAK	.150**	1					
CLEAN	.312**	.166**	1				
REST	.156**	.161**	.171**	1			
EMOTION	.193**	.136*	.228**	.308**	1		
SMILE	-0.029	.202**	.131*	0.092	.213**	1	
STUDY	.274**	0.105	0.083	.202**	.331**	0.052	1
SOCIAL	0.056	.195**	.207**	.145*	.382**	.367**	.188**

\*:P<.05, \*\*:P<.01, \*\*\*:P<.001

### 3.4.1 형성적 측정구조로 COIDP 분석

COIDP 구강건강 관련 삶의 질 도구의 형성적 측정구조를 분석하기 위하여 유사한 구성개념인 COHIP를 포함하는 구조모형을 구성하여 추정을 실시하였다

(그림 28).

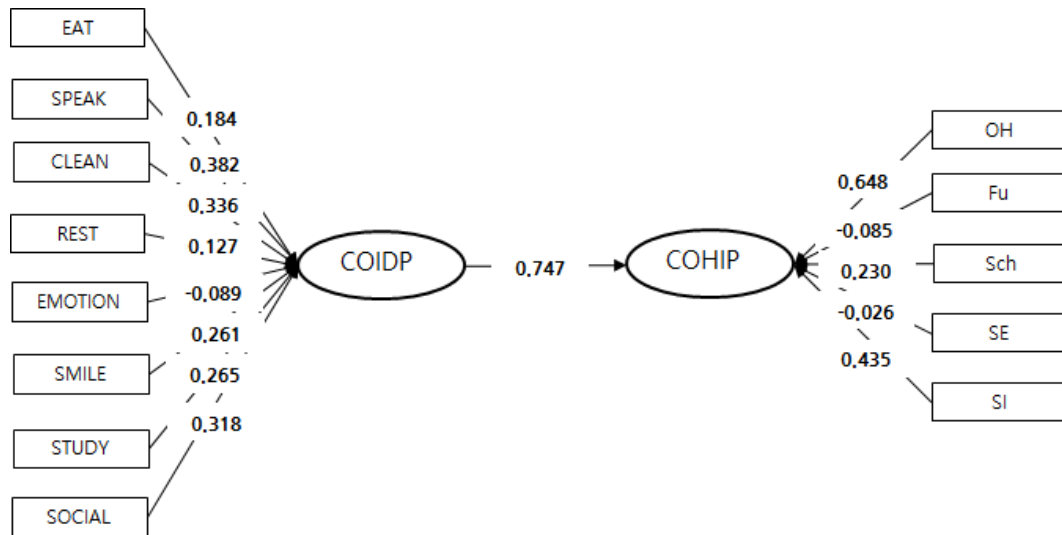


그림 28. COIDP 도구의 형성적 측정모형(COHIP 구성개념 포함)

OH: Oral Health, Fu: Functional Well-being, Sch: School activities, SE: Social/Emotional Well-being, and SI: Self-image.

COIDP의 형성적 측정모형의 추정 결과 외부가중치 중 쉬기, 웃기 및 친구 만나기 등의 외부가중치가 0.05 유의수준에서 비유의한 것으로 평가되었고 다중 공선성의 문제는 발생하지 않는 것으로 나타났다(표 27). 형성적 측정모형의 추정 에 있어서 PLS-SEM과 PCA 방법에 의한 가중치의 추정치 차이는 매우 크게 나타나 여덟 개 지표변수 중 여섯 개가 0.1에 근접하거나 그 이상의 큰 차이를 보였다(표 27).

표 27. COIDP의 형성적 측정모형의 질 평가 (n=306)

Indicators	가중치	PLS-SEM		주성분 분석*
		P-value	VIF	가중치
Eating	0.32	0.001	1.3	<b>0.23</b>
Speaking	0.26	0.031	1.8	<b>0.17</b>
Cleaning	0.18	0.020	1.3	0.20
Resting	0.34	0.053	1.9	0.28
Emotional Stability	0.38	0.010	1.4	<b>0.22</b>
Smiling	0.13	0.364	1.5	<b>0.25</b>
Studying	0.27	0.004	1.1	<b>0.13</b>
Social activity	-0.089	0.458	1.8	<b>0.28</b>
Mean	0.22			0.22

\*: 동일 자료에 대하여 1요인을 적용한 주성분 분석 결과

COIDP 구성개념과 COHIP 구성개념 간의 구조계수는 0.75로 유의한 값을 보였다(표 28). 만약 이 모형을 일종의 중복 분석으로 간주한다면 결정계수 0.558은 수렴타당도 면에서 만족하지 못하는 것으로 볼 수 있다(표 29).

표 28. COIDP의 형성적 측정모형과 COHIP(n=306)

	Structural Coefficients	
	Estimate	P-value
COIDP	0.747	<.001

COIDP의 R-square=0.558

표 29. COIDP의 형성적 측정모형 및 구조모형의 평가

	평가 항목	평가 기준	COIDP 모형 평가
형성적 측정모형	수렴타당도	중복 분석: 수정된 R <sup>2</sup> 값 이 0.81, 0.64 초과	0.558이므로 기준치 이하
	다중공선성	VIF가 5 초과 / 공차가 0.2 미만	모든 지표변수가 조건에 만족함
	외부 가중치 타당	외부가중치 유의	세 지표변수가 0.05 유의

	성		수준에서 유의하지 않음
구조모형	결정계수 ( $R^2$ )	small-0.02, medium-0.13, large-0.26	0.558로 큰 편임

VIF: Variance Inflation Factor.

## 2) COIDP 도구에 대한 반영적 측정모형 적용

비교의 목적으로 COIDP 도구를 반영적 측정모형으로도 분석하기 위하여 COHIP를 유사한 구성개념으로 도입하여 구조모형을 구성하였다(그림 29).

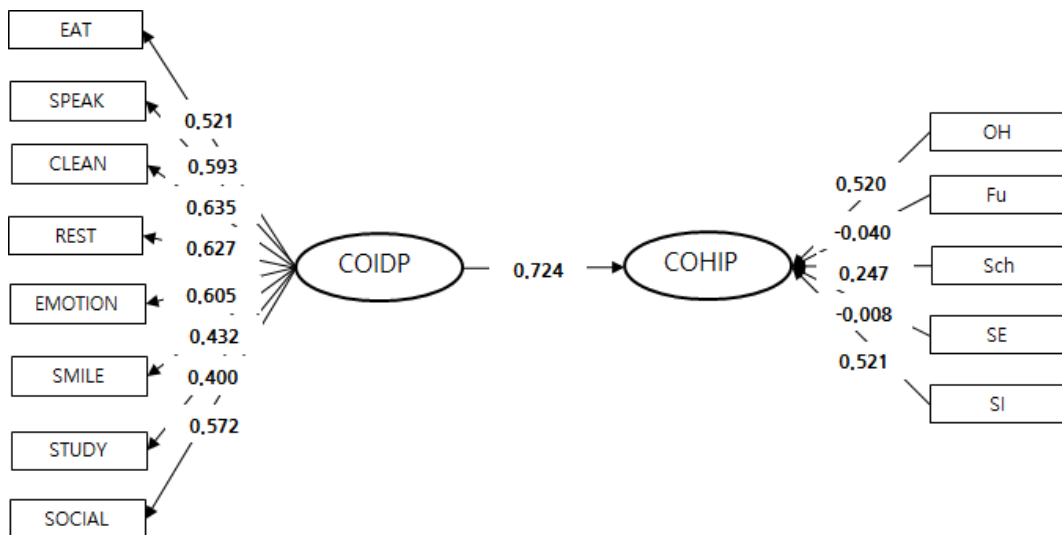


그림 29. COIDP 도구의 반영적 측정모형(COHIP 구성개념 포함)(n=306)

OH: Oral Health, Fu: Functional Well-being, Sch: School activities, SE: Social/Emotional Well-being, and SI: Self-image.

표 30은 COIDP 도구의 반영적 측정모형의 구성개념에 대한 PLS-SEM에 의한 분석을 실시한 후, 질 평가 사항을 정리한 것이다. 지표변수 신뢰성 면에서

는 모든 지표변수들의 요인계수가 0.40~0.64로 적정한 기준인 0.7 보다 작았고 평균추출분산(AVE)은 0.31로 작았다. 내적 일관성 면에서 복합신뢰도(Composite Reliability, CR)는 0.78 로 높았다(Nunally & Bernstein, 1994).

표 30. COIDP의 반영적 1차 구조의 질 평가 기준(n=306)

Indicators	요인계수	AVE	복합신뢰도
Eating	0.57	0.31	0.78
Speaking	0.43		
Cleaning	0.52		
Resting	0.64		
Emotional Stability	0.59		
Smiling	0.63		
Studying	0.40		
Social activity	0.61		

COIDP 구성개념과 COHIP 구성개념 간의 구조계수는 0.72로 유의한 값을 보였다(표 30). 만약 이 모형을 일종의 중복분석으로 간주한다면 결정계수 0.525는 수렴타당도 면에서 만족하지 못하는 것으로 볼 수 있다(표 31).

표 31. 반영적 COIDP 모형과 COHIP(n=306)

	Structural Coefficients	
	Estimate	P-value
COIDP	0.724	<.001

COIDP의 R-square=0.525

COHIP 도구의 이차구조의 결정계수는 형성적 이차 관계의 특성 상 1.0으로 나타났고 전체 구조모형의 적합도를 평가하는 GoF는 0.833 으로 큰 GoF의 기준인 0.36을 크게 상회하므로 적합도가 매우 좋다고 평가할 수 있다.

$$GoF = \sqrt{AVER^2} = \sqrt{0.694*1} = 0.833$$

표 32에는 COHIP 2차 계층모형의 PLS-SEM 추정과 관련한 평가결과를

요약하였다.

표 32. 반영적 COIDP 측정모형과 구조모형 평가

평가 단계	평가 사항	평가 기준	SF-35 모형 평가
반영적 측정모형 평가	내적 일관성 / 구성개념 신뢰성 지표변수	CR 0.6~0.7 이상이면 수용가능. 0.9초과 시 주의	0.78이므로 만족함
	신뢰성 지표변수	외부 요인계수 0.7 이상	모든 지표변수가 기준에 미달함
구조모형 평가	수렴타당도	AVE 0.5이상	0.31로 매우 미달함
구조모형 평가	결정계수 ( $R^2$ )	$R^2$ : small-0.02, medium-0.13, large-0.26	0.525 로 큼.
	Global goodness of fit	$GoF = \sqrt{AVE R^2}$ , small-0.1, medium-0.25, large-0.36	0.40으로 매우 만족.

CR: Composite Reliability, AVE: Average Variance Extracted, GoF: Goodness of Fit.

### 3) COIDP 도구에 대한 형성적 및 반영적 측정모형 비교

COIDP 도구에 대한 형성적 및 반영적 측정모형 적용 결과의 가중치와 요인계수가 각각 계산된 바 있다. 추정된 구조계수는 0.747 및 0.724로 크기가 비슷하였다 그러나 COIDP와 구조적 관계로 설정된 COHIP는 동일한 형성적 관계로 적용되어 있다. 그러나 COIDP에 대한 형성적 및 반영적 측정모형 적용에 따른 가중치의 크기를 비교하면 Functional 차원에서 0.13, Self Image 차원에서 0.08의 가중치 간 차이를 보이는 것으로 나타났다(표 33).

표 33. COIDP 모형화에 사용된 COHIP 측정모형의 가중치 비교

	형성적 COIDP 측정모형(그림 29)	반영적 COIDP 측정모형
Functional	0.65	<b>0.52</b>
Oral Health	-0.09	-0.04
School Environment	0.23	0.25
Social-Emotional	-0.03	0.01
Self Image	0.44	<b>0.52</b>
	0.24	0.25

### 3.5 COHIP 모형에 대한 교차타당화(Cross-validation)

표본이 크고 형성적 및 반영적 측정구조가 동시에 포함된 COHIP 자료를 선정하여 교차타당도를 검증하였다. 총 2,068명의 조사 자료를 확률적으로 대략 반으로 나누어 각각을 PLS-SEM 모형으로 분석하여 외부가중치, 요인계수 및 구조계수 등의 크기를 비교하였다(표 34, 표 35, 표 36). 비교 결과 0.1을 상회하는 차이는 발견되지 않았고 대체로 유사하였다. PLS-SEM모형에서는 PLS-SEM에서와 같이 적합도 지표를 이용하여 요인계수의 동일성을 통계적으로 검증할 수 있는 방법이 없으므로 단순히 추정치 차이의 크기에 의존하여 평가하였다.

표 34. COHIP의 형성적 구조계수 비교(반복지표변수 모형, Mode B)

	Sample 1 (n=1011)		Sample 2 (n=1057)		Difference
	Estimate	P-value	Estimate	P-value	
Oral Health	0.362	<.001	0.349	<.001	0.013
Function-School	0.367	<.001	0.380	<.001	-0.013
Socio-Emotional	0.359	<.001	0.369	<.001	-0.010
Self-Image	0.193	<.001	0.192	<.001	0.001
Mean	0.321		0.323		-0.002

COHIP의 R-square=1.00

표 35. COHIP의 형성적 1차 구조 'Oral Health'에 대한 요인계수 비교

Indicators	Sample 1 (n=1011)		Sample 2 (n=1057)		Difference
	가중치	P-value	가중치	P-value	
OH1	0.37	<.001	0.35	<.001	0.02
OH2	-0.04	0.471	0.04	0.406	-0.08
OH3	0.17	0.001	0.25	<.001	-0.08
OH4	0.20	<.001	0.22	<.001	-0.02
OH5	0.10	0.036	0.11	0.032	-0.01
OH6	0.19	<.001	0.22	<.001	-0.03
OH7	0.20	<.001	0.18	<.001	0.02
OH8	0.15	0.013	0.22	<.001	-0.07
OH9	0.16	0.004	0.07	0.242	0.09
OH10	0.20	0.001	0.13	0.008	0.07
Mean	0.17		0.18		-0.009

표 36. COHIP의 반영적 1차 구조의 요인계수 비교

Domain	Indicators	Sample 1	Sample 2	Difference
		(n=1011) 요인계수	(n=1057) 요인계수	
Functional-School	Fu1	0.62	0.58	0.04
	Fu2	0.67	0.68	-0.01
	Fu3	0.62	0.71	-0.09
	Fu4	0.63	0.65	-0.02
	Fu5	0.62	0.68	-0.06
	Fu6	0.54	0.57	-0.03
	Sch2	0.68	0.69	-0.01
	Sch3	0.66	0.68	-0.02
Social-Emotional	Sch4	0.53	0.62	-0.09
	SE1	0.77	0.77	0.00
	SE2	0.70	0.69	0.01
	SE3	0.76	0.76	0.00
	SE4	0.71	0.72	-0.01
	SE5	0.75	0.74	0.0
	SE6	0.82	0.82	0.00
	SE7	0.60	0.58	0.02
Self Image	SE8	0.68	0.69	-0.01
	SI3	0.77	0.80	-0.03

	SI4	0.74	0.71	0.03
	SI5	0.82	0.83	-0.01
	SI6	0.69	0.68	0.01
Mean		0.68	0.70	-0.01

#### 4. 기존 방법과 모형에 의한 구성개념 점수 비교

##### 4.1 SF-36

다양한 방법에 의한 SF-36의 통합 점수를 산출하고 비교하였다. 표 37에는 통상적으로 사용하는 8개의 차원 점수 평균, 35개의 문항 점수 평균 및 PLS-SEM모형에 의한 가중치에 의한 통합 점수의 평균값과 표준편차를 나타내었다. 세 방법에 의한 차이는 매우 미미한 것으로 나타났다. 또한 세 방법에 의한 점수 간 상관을 확인한 결과 피어슨 상관계수가 0.99를 초과하여 매우 유사성이 높은 것으로 사료되었다(표38).

표 37. SF-36 차원, 문항 점수 평균 및 가중치에 의한 통합(n=188)

	Mean (dimension scores)	Mean (35 item scores)	Weighted Mean (by PLS model)*
Mean	80.23	80.73	80.24
SD	11.99	11.62	11.81
Min	39.58	44.28	43.28
Max	100	100	100

$$*Weighted\ mean = \frac{\sum_i W_i Score_i}{\sum_i W_i}$$

표 38. 여러 방법에 의한 SF-36 통합점수의 피어슨 상관계수(n=188)

	Mean (dimension scores)	Mean (35 item scores)	Weighted Mean (by PLS model)
Mean (dimension scores)	1		
Mean	0.977***	1	

(35 item scores)			
Weighted Mean (by PLS model)	0.991***	0.996***	1

\*\*\*:  $p < .001$

## 4.2 COHIP

네 개의 차원으로 구성된 COHIP의 통합 점수를 4개의 차원 점수 평균, 31개의 문항 점수 평균 및 PLS-SEM모형에 의해 산출한 가중치에 의한 통합 점수 등 세 가지 방법으로 구하였다. 표 39와 표 40에 따르면 평균값과 표준편차 등 산술적 수치의 차이는 매우 미미하였고 세 방법에 의한 점수 간 이변량 상관을 확인한 결과 피어슨 상관계수가 0.97를 초과하여 매우 유사성이 높았다.

표 39. COHIP 차원, 문항 평균 및 가중치에 의한 통합(n=2068)

	Mean (dimension scores)	Weighted Mean (by PLS model)*	Mean(31 item scores)
Mean	3.05	3.18	3.18
SD	0.43	0.41	0.41
Min	1.28	1.33	1.35
Max	4.00	4.00	4.00

표 40. 여러 방법에 의한 COHIP 통합 점수의 피어슨 상관계수(n=2068)

	Mean (dimension scores)	Mean (31 item scores)	Weighted Mean (by PLS model)
Mean(dimension scores)	1		
Mean(31 item scores)	0.975***	1	
Weighted Mean (by PLS model)	0.983***	0.998***	1

\*\*\*:  $p < .001$

### 4.3 EQ-5D

EQ-5D는 일반적으로 인구집단의 대표성 있는 집단에서 선호도 조사를 실시하고 그 결과에 의거한 선호도기반 점수를 통합점으로 사용하여 왔다. 우리나라에서도 Jo 등(2008)의 연구와 강은정 등(2006)의 연구에서 조사된 선호도 점수가 보고된 바 있다. 이에 두 연구에서 보고한 점수와 PLS-SEM의 형성적 측정 모형 적용 결과 도출된 각 항목 당 가중치로 계산된 통합점수를 비교하였다. 더불어 많은 경우 연구자들이 사용하는 합산점수 역시 비교의 대상이 된다.

표 41에는 3 수준의 응답지를 가지는 한국어판 EQ-5D 도구에 대한 Jo et al(2008)과 강은정 등(2006)에 의하여 각각 조사된 EQ-5D 개별 수준에 대한 선호도 기반 가치평가 점수가 수록되어 있다. 또한 PLS-SEM 모형을 통하여 추정된 가중치를 이용하여 해당 수준의 점수 추정치를 계산한 결과를 비교하고 있다. 이때 원 설문이 좋은 건강에 낮은 점수를 부여하고 있으므로 역코딩하여 적용하였다.

표 41. 선호도기반 및 PLS-SEM 추정에 의한 EQ-5D의 점수

EQ-5D	Preference-based		PLS <sup>3)</sup>	EQ-5D	Preference-based		PLS
	Jo 등 <sup>1)</sup>	Kang <sup>2)</sup>			JMO	Kang	
11112	0.922	0.7358	2.746	22222	0.587	0.4155	2.000
11113	0.8	0.4154	2.492	22233	0.358		1.422
11121	0.91	0.7681	2.676	22323	0.252		1.621
11122	0.812		2.422	22331	0.478		1.805
11131	0.778	0.3732	2.352	23232	0.34	0.155	1.586
11133	0.731	0.1728	1.844	23313	0.107		1.854
11211	0.906	0.7862	2.875	23321	0.295		2.039
11312	0.705	0.3362	2.495	32211	0.598	0.0731	2.371

12111	0.908	0.8113	2.910	32223	0.135	-0.0782	1.539
12121	0.798		2.586	32232	0.1		1.470
12211	0.797		2.785	32313	0.329	-0.0658	1.738
12222	0.665		2.207	32331	0.394		1.599
12223	0.581		1.952	33212	0.263		2.027
13212	0.625		2.440	33232	0.203		1.379
13311	0.576	0.2364	2.569	33321	0.118		1.832
13332	0.226		1.667	33323	-0.161	-0.148	1.324
21111	0.902	0.8161	2.793	33333	-0.708	-0.163	1.000
21133	0.701		1.637				
21222	0.759		2.090				
21232	0.584		1.766				
21312	0.68		2.289				
21323	0.545		1.711				
22112	0.751		2.449				
22121	0.781		2.379				
22122	0.651		2.125				

1) Jo 등, 2008; 2) 강은정 등, 2006; 3) Weighted mean =  $\frac{\sum_i W_i \text{Score}_i}{\sum_i W_i}$

상기한 방법 이외에 일반 연구자들이 손쉽게 사용하고 있는 합산점수를 포함하여 네 가지 서로 다른 방법에 의해 도출된 통합점수 간 이변량 상관관계를 확인한 결과 스피어만 상관계수가 0.84~0.96의 범위로 나타나 대체로 점수 간의 유사성이 높았다(표 42).

표 42. EQ-5D의 수준별 점수 간의 스피어만 상관계수

	N	선호도 기반		항목 점수합	PLS weight
		조민우 <sup>1)</sup>	강은정 <sup>2)</sup>		
조민우 등, 2008	42	1			
강은정 등, 2006	17	0.892***	1		
항목 점수합	42	0.962***	0.920***	1	
PLS weight	42	0.843***	0.902***	0.937***	1

1) Jo 등, 2008; 2) 강은정 등, 2006; \*\*\*:  $p < .001$

## V. 논 의

### 1. 결과해석 및 의의

복잡한 다차원적 구조를 가지는 HRQoL 측정도구의 내부 구조(structure)나 구성타당도(construct validity)를 규명하기 위하여 많은 연구가 진행되어 왔다(McHorney et al, 1993; Cho et al, 2016). 또한 여러 차원 간을 통합하는 요약점수를 얻는 방법 역시 다양하게 고안되고 있다(강은정 등, 2006).

HRQoL 도구의 다차원적 하부개념들은 반영적 또는 형성적 측정구조를 가지며 하부개념들을 종합하는 HRQoL은 형성적 관계를 가지는 것으로 보인다. 그러나 기존 연구에서 HRQoL 도구들은 심리측정적 전통을 따라 반영적인 구조와 관계로만 설정되어 연구되고 있다.

본 연구에서는 대표적인 HRQoL 도구인 SF-36, EQ-5D, COHIP 및 COIDP의 측정구조를 심층적으로 고찰하여 적절한 구조방정식 모형으로 도출한 후 실증적인 자료를 이용하여 이를 PLS-SEM을 적용하여 분석하였다. 더불어 그 결과로 산출된 가중치를 활용하여 통합된 HRQoL 요약점수를 제시하고 기존 요약점수 산출방법과 비교하였다. 대체로 HRQoL 측정도구에 대한 구조적 관점 및 차원 간의 통합에 초점을 맞추어 이 연구를 수행하였다.

첫째, 다양한 HRQoL 측정도구의 구조를 반영적 및 형성적 특성을 중심으로 고찰하였다. HRQoL 측정도구의 구조에 대한 연구들은 전통적으로 심리측정적 방법론에 따라 이루어져 왔으며, 자료 분석에는 CB-SEM 모형이 주로 활용되었다. 따라서 HRQoL 도구의 구조와 관련한 기존 연구에서는 오로지 반

영적 측정모형과 반영적 2차 관계를 구현하여 구성타당도를 검증하는 것이 일반적으로 이루어지고 있다. 그러나 HRQoL 측정도구에 존재하는 일차 구성개념, 즉 차원들과 이차 구성개념인 HRQoL 간의 관계는 내용 상 반영적 이라기 보다는 형성적인 구조로 보는 것이 적절하다(McHorney, 1993). 더구나 측정모형 자체가 형성적인 특성으로 이루어진 경우도 존재한다.

따라서 본 연구에서는 먼저 다양한 HRQoL 측정도구를 내용적으로 검토하고 각 차원의 측정구조 및 차원들과 상위 구성개념과의 관계가 반영적인지 형성적인지를 고찰함으로써 HRQoL 측정도구 각각에 적합한 구조방정식 모형을 수립하여 이를 이후 실증 분석의 기초 모형으로 사용하였다.

고찰 결과, SF-36 도구는 반영적 측정구조를 가지는 8개의 차원 및 이들을 형성적 관계로 연결하는 한 개의 이차 구성개념을 가지는 이차 계층적 구조모형으로 설정하였다. EQ-5D 도구는 다섯 개의 서로 다른 측면에 대한 지표변수가 형성적 측정모형으로 하나의 구성개념으로 이루는 구조를 설정하였다. COHIP 도구는 총 4개의 차원으로 설정하였는데 그 중 'Oral Health'는 형성적 측정모형으로 나머지 차원들은 반영적 측정모형으로 설정하고 각 차원들은 형성적 관계를 가지는 하나의 이차 구성개념으로 연결되는 이차 계층적 구조모형으로 설정하였다. COIDP 도구는 여덟 개의 서로 다른 측면에 대한 지표변수가 형성적 측정모형으로 하나의 구성개념을 이루는 구조를 설정하였다.

둘째, 반영적 측정모형과 형성적 측정모형을 용이하게 수용할 수 있다고 여겨지는 PLS-SEM을 실증분석에 적용하는 것이 적절하다는 판단 하에서 반영적 성격의 자료에 대한 PLS-SEM의 추정 오류 정도를 평가하고자 하였다. 구체적으로 모의실험을 통하여 형성적 측정모형의 원리에 의하여 복합지수를 기반으로 추정하는 PLS-SEM을 반영적 특성을 가진 자료에 적용하였을 때 발생하는 추정 오류의 정도를 평가하였다.

가상 데이터는 반영적 구조를 가지는 데이터로서 요인계수(0.4, 0.6,

0.8), 요인 당 문항 수(3, 6, 12) 및 표본크기(150, 300, 500, 1000) 등의 조건을 조합하여 총 36개의 서로 다른 경우를 가정하였다. 36개의 조건 당 각각 1000개의 데이터를 생성하여 요인분석법 및 주성분분석법을 반복 적용하여 분석하고 그 결과를 비교하였다. 모의실험 결과에 의하면, 요인별 지표변수의 수가 3개인 모든 경우와 요인계수가 0.4로 낮으면서 요인별 지표변수의 수가 6개 이하인 경우 추정된 요인계수의 크기가 요인계수 참값보다 0.1이상 크게 추정되는 심각한 오류가 나타났다. 이는 Widaman(1993)의 연구와 일치하는 결과이다.

이 결과를 참고하여 실증 분석 대상인 SF-36 및 COHIP 도구의 반영적 측정구조의 추정에 PLS-SEM 모형을 적용할 때 예상할 수 있는 요인계수의 추정오류를 가늠해 볼 수 있다. SF-36은 요인계수 평균이 0.72, 차원 당 평균 지표변수 4.38개이고 COHIP는 요인계수 평균이 0.61, 차원 당 평균 지표변수 7개여서 추정 요인계수와 참값의 차이가 대략 0.08이하로 심각한 편향이 없을 것으로 예상하였다. 반영적 지표변수의 요인계수는 일반적으로 상당히 높아서 PLS-SEM을 적용할 때 추정 오류가 심각하지 않을 수 있다. 그러나 CB-SEM에서 보통 요인계수가 0.4이상인 경우를 유효한 지표변수의 판단기준으로 삼고 있으므로 요인계수가 낮은 경우 큰 편향이 실제로 나타날 수 있으므로 주의하여야 한다.

셋째, PLS-SEM을 이용하여 HRQoL 측정도구에 대한 실증분석을 실시한 결과 모형이 잘 적합되었다. 상기한 HRQoL 도구에 대한 PLS-SEM 적용 결과 반영적 측정모형은 내적일관성, 지표변수 신뢰성, 수렴타당성 및 판별 타당성이 대부분 충족되었고 형성적 측정모형을 가지는 차원 역시 다중공선성 및 외부 가중치 타당성 면에서 적절하였다. 일차 구성개념들이 형성적 관계로 정의된 모형으로 이차 계층 구조를 포함한 전체 구조방정식에도 PLS-SEM이 잘 적용되었다.

SF-36과 COHIP 도구에 대한 모형 적합 결과 반영적 측정모형은 내적 일관성, 지표변수 신뢰성, 수렴타당성 및 판별 타당성이 충족되었다. COHIP 도구의 형성적 측정모형인 Oral Health 차원 역시 다중공선성 및 외부 가중치 타당성 면에서 모형이 적절하였다. EQ-5D 도구에 대한 형성적 모형 측정은 다중공선성, 외부 가중치 타당성 및 결정계수 등의 평가항목에서 적절한 것으로 나타났다. 반면 COIDP 도구의 경우 외부 가중치 타당성 기준에 미달하는 변수가 여덟 개 중 세 개였다.

예상한 바와 같이 반영적 측정모형의 자료에 대하여 PLS-SEM을 적용한 결과의 추정 요인계수는 CB-SEM의 적용 시와 비교하여 상당한 차이를 보였는데 특히 요인계수가 작은 경우 오차분산이 커지므로 두 방법의 차이가 컸다. EQ-5D 도구와 COIDP 도구의 자료에 대하여 PLS-SEM과 PCA로 계산된 가중치는 서로 상당한 차이를 보였으나 COHIP의 'Oral Health'차원에서 추정된 가중치의 경우에는 대체로 유사하였다.

PLS-SEM과 PCA의 적용 결과의 이러한 차이는 형성적 측정구조에 연결된 다른 변수의 영향력을 반영하는 것으로 EQ-5D나 COIDP의 형성적 측정구조에 EQ-VAS 나 COHIP가 연결되었는지의 여부에 따른 차이이다. 구성개념의 측정구조 내 가중치 또는 요인계수가 연관된 변수의 영향으로 크게 변화한다면 형성적 측정구조가 독립적인 측정모형으로 유효한 지에 의문이 제기될 수도 있다. 또한 가중치가 다른 변수와의 관계에 따라 변화한다면 가중치의 일반화에 혼돈을 초래하여 어떤 가중치를 사용하여 복합점수를 산출하는 것이 적절한 지 판단하여야 하는 문제가 발생한다.

넷째, 여러 차원들을 다양한 방법으로 통합하여 HRQoL 요약점수로 표현하여 비교하였다. SF-36 및 COHIP 도구의 경우 PLS-SEM모형에서 산출된 차원 간 통합 가중치를 이용한 경우 일반적으로 요약점수로 사용되고 있는 차원점수 평균이나 전체 문항점수 평균과의 상관 정도가 매우 높은 것으로 나타

났다. PLS-SEM 적용 결과 산출된 가중치로 계산한 요약점수는 EQ-5D 도구의 선호도 기반 점수와 비교하였을 때에도 상당한 정도로 유사하였다. 따라서 본 연구의 결과에 따르면 사용자의 편리성 면에서 유리한 합산 점수나 평균 점수를 구성개념의 요약점수로 사용하는 것이 대체로 적절하다고 사료되었다.

본 연구에서는 대표적인 HRQoL 도구들을 분석함에 있어서 전체 측정도구의 구조를 아우르는 통계적 모형을 적용하였다. 통계적 모형 적용의 장점은 다음과 같다.

첫째, 측정모형의 타당성 평가가 가능하다. 모형 적용 후 분석 결과에 대하여 다양한 신뢰도 및 타당도 검증방법을 적용함으로써 전체 측정구조의 적합도, 개별 변수의 신뢰도 및 구성개념의 타당도 등을 평가할 수 있다.

둘째, 다양한 표본에 대한 분석 결과를 비교하고 평가할 수 있고 교차타당도 등의 검증을 통하여 측정모형의 일반화가 가능하다는 장점이 있다. 예를 들면 선호가중치의 경우 대표성 있는 표본에 대한 일회적인 조사로 가중치를 확정하고 유사한 인구집단에 대하여 일괄적으로 적용하는데, 이 가중치에 대한 신뢰성이나 타당성에 대한 평가방법은 존재하지 않는다. 더불어 서로 다른 조사에서 상이한 가중치가 얻어지는 경우 이를 비교 평가하는 기준이 마련되어 있지 않다.

통계적 모형의 적용을 통하여 서로 다른 표본을 사용하여 수립된 측정모형의 요인계수나 가중치 등의 결과에 대하여 교차타당화(cross-validation)를 시도할 수 있다. 본 연구의 결과에서 자료를 반분하여 동일한 모형을 적용한 결과 유사한 추정치를 얻을 수 있었다. 그러나 PLS-SEM에서는 CB-SEM에서와 같이 모형에 의하여 재생된 관계가 원데이터에 적합한 지에 대한 평가 방법이 부재하므로 그 결과를 요약된 통계치로 제시할 수는 없다는 한계가 있다. 교차타당화를 시도하고 상이한 분석 결과 간 차이의 원인을 규명하는 과정은 측정모형의 일반화를 위한 노력의 출발점이 될 수 있다.

본 연구에서는 SmartPLS 소프트웨어를 이용하여 PLS-SEM 모형을 HRQoL 측정도구에 적용하였는데 측정모형 수립이라는 관점에서 볼 때 그 과정에서 나타난 몇 가지 의문점 또는 한계점을 다음과 같이 제기할 수 있다.

첫째, 모형에 함께 포함된 다른 변수로 인하여 측정모형 내 추정치의 큰 변화가 나타난다는 점이다. PLS-SEM 모형은 1차 형성적 측정모형 자체로는 식별이 불가능하므로 이 연구에서는 EQ-5D는 EQ-VAS를, COIDP는 COHIP를 포함하여 모형을 구성하였다. 형성적 측정모형인 COHIP를 구성하는 가중치는 연관된 COIDP가 형성적으로 정의되었는지 반영적으로 정의되었는지에 따라 변동하였다(표 33, 143쪽). 그렇다면 COHIP의 어떤 가중치가 보다 적절한 가중치인지, 그에 대한 평가 기준이 있는지 등이 의문시 된다. 과연 모형에 포함된 다른 변수의 변화에 따라 측정모형 자체가 일관된 결과를 산출하지 못한다면 한 구성개념의 독자적인 측정모형으로 보기에 큰 문제가 있다. 반면 CB-SEM 모형은 일반적으로 측정모형의 추정치가 외부의 다른 구조 변수의 유무와 무관하게 거의 미미하게 변동하여 매우 안정적이어서 측정모형 그 자체에 충실한 결과를 제공한다.

둘째, 반영적 측정모형의 경우, 수렴타당도 기준이 AVE 0.5인데 이는 전체 분산의 50% 만을 공통성으로 하는 경우여서 나머지 50%의 분산은 구성개념과 무관한 고유분산에 해당하는 경우이다. 이런 경우에도 지표변수의 전체 변동을 포함하는 복합지수(composite)를 구성개념 값으로 이용하는 것은 상호 모순적일 수 있다.

셋째, 형성적 모형의 수렴타당성을 평가하는데 중복 분석이 제안되었는데 EQ-5D나 COIDP가 모두 수렴타당의 조건에 충족하지 못하였다. 현실적으로 서로 다른 방법으로 조사한 유사한 구성개념간의 구조계수가 0.8을 상회하기는 쉽지 않을 것으로 보여 중복 분석의 실효성에 의문을 제기할 수 있다.

넷째, SF-36과 COHIP와 같이 이차 계층구조인 도구에서 형성적 관계로

생성되는 이차 구성개념을 가지는 경우, 모형의 결정계수는 대체로 1이다. 그 이유는 이차 구성개념의 복합 점수는 반복되는 지표변수에 의한 차원들에 의하여 완벽하게 설명되기 때문이다. 이런 경우, 모형의 예측력이나 적합도를 평가하는 데 결정계수를 사용하기 어렵다. 또한 Smart PLS의 기술적 문제일 수 있는데, 이차계층구조에서는 f 제공이 허용 범위를 넘어서는 값이 표기된 사례가 있었다.

다섯째, PLS-SEM의 경우 반영적 측정모형의 적합 시 지표변수의 특수분산 간 상관의 설정이 불가능하다. 이론적으로 유사한 방법론이나 공통요인과 무관한 내용의 유사성 등으로 인한 지표변수의 특수분산 간 상관이 발생할 수 있으며 CB-SEM에서는 이를 모형에 반영하고 있다. 그러나 PLS-SEM에서는 특수분산 간 상관의 설정이 가능하지 않다.

본 연구의 제한점은 다음과 같다.

첫째, PLS-SEM이 심리측정 분야에서나 보건학 분야에서 활용 사례가 매우 제한된 상태에서 연구가 진행된 면이 있으므로 이후 다양한 PLS-SEM 적용 사례를 확보하여 보다 풍부하고 심도 있는 고찰이 필요하다. 둘째, 반영적 및 형성적 측정모형의 구별이 이론적 내용적인 면에서 이루어져 주관성이 개입될 우려가 있으므로 보다 객관적인 판단 기준 개발의 필요성이 있다. 셋째, 본 연구에서 다루었던 4개의 HRQoL 도구 이외에 다른 다양한 구성개념 측정도구들에 관하여 PLS-SEM 모형을 적용하여 일반화 할 필요성이 있다.

요약하면 본 연구의 가장 큰 의의는 전통적으로 반영적 구조에만 의거하여 측정구조를 분석하던 여러 HRQoL 측정도구에 대하여 형성적 및 반영적 관계를 적절하게 규정하고 올바르게 정의된 측정모형에 PLS-SEM을 적용하여 경험적으로 구현하였으며 구현된 모형의 결과를 기반으로 다차원적 HRQoL 측정도구의 차원을 통합하는 방법을 제안하였다는 점에 있다.

## 2. 후속연구 제안

후속연구에서는 연구범위가 다양한 구성개념들로 확장되어야 할 필요성이 있고 또한 다양한 조건으로 모수를 미리 정하고 가상데이터를 생성하여 두 방법의 수행을 평가할 필요성이 있다.

첫째, 보건학 분야의 다양한 구성개념으로 그 범위를 확대하여 그들의 구조를 이론적으로 또한 경험적 데이터 분석을 통하여 밝히고 각 구성개념들의 올바른 운용방법에 관하여 연구할 필요성이 있다. 본 연구에서 다루었던 4개의 HRQoL 도구 이외에 다른 다양한 구성개념 측정도구들에 관하여 PLS-SEM 모형을 적용하여 반영적 및 형성적 측정구조의 개념을 확립하고 일반화할 필요성이 있다.

둘째, 이 연구의 모의실험에서는 반영적 구조의 데이터를 대상으로 하였는데, 형성적 구조의 데이터나 이차 계층구조 등 다양한 조건과 구조를 설정하고 데이터를 생성하여 분석함으로써 사전에 정한 모수의 조건과 반영적 및 형성적 측정모형을 이용한 방법 간의 비교가 필요하고 또한 CB-SEM 및 PLS-SEM 적용 결과를 비교하고 그 수행을 평가하는 연구가 필요하다.

셋째, 현재 반영적 및 형성적 측정모형의 구별이 이론적 내용적인 면에서 한정되어 다소 주관적인 판단이 개입될 우려가 있으므로 이를 객관화시킬 수 있는 방법론의 개발이 요망된다. 이론적으로 보다 정교한 판단 기준을 마련하거나 합리성 있는 계량적인 판단 기준을 개발할 필요성이 있다.

넷째, PLS-SEM 전체 모형의 적합도 평가 방법에 관한 후속연구가 필요하다. CB-SEM에서는 원자료의 상관행렬과 모형 적용 결과 재생된 상관행렬 간의 차이를 기초로 하여 다양한 적합도 지수를 개발하고 적용하고 있는데 PLS-SEM에서는 결정계수와 AVE에 의해서만 전체 모형의 적합도를 평가하고 있다. PLS-SEM의 추정 결과에 대해서도 원자료의 재현성을 평가할 수 있는

보다 근본적인 의미의 적합도 평가가 필요한 것으로 사료된다.

### 3. 종합논의 및 결론

본 연구는 HRQoL 측정도구의 이론적 구조를 반영적 및 형성적 측정모형과 관련하여 고찰하고 적절한 구조방정식모형을 규명하였으며 이 모형을 토대로 실제 데이터 분석을 통하여 이론적 모형을 경험적으로 구현하였다. HRQoL 도구는 다차원적인 도구로서 각각의 하부개념들은 반영적 또는 형성적 측정구조를 가질 수 있고 그러한 하부개념들을 종합하는 이차 구성개념인 HRQoL은 하부개념들과 형성적 관계를 가지는 것으로 보인다. 그러나 기존 연구에서 HRQoL 도구들은 심리측정적 전통을 따라 반영적인 구조와 관계로만 설정되어 연구되고 있다. 본 연구에서는 건강 및 구강건강관련 삶의 질 도구로 대표적인 SF-36, EQ5D, COHIP, 및 OIDP의 측정구조를 심층적으로 고찰하고 실증적인 자료를 이용하여 PLS-SEM을 적용하였다.

첫째, 본 연구에서는 다양한 HRQoL 측정도구를 내용적으로 검토하고 각 차원의 측정구조 및 차원들과 상위 개념과의 관계가 반영적인지 형성적인지를 고찰함으로써 HRQoL 측정도구 각각에 적합한 구조방정식모형을 수립하여 이를 실증 분석의 기초 모형으로 사용하였다.

둘째, 형성적 측정모형의 원리에 의하여 복합지수를 기반으로 추정하는 PLS-SEM을 반영적 특성을 가진 자료에 적용하였을 때 발생하는 추정 오류의 정도를 평가하고자 하였다. 가상 데이터 생성 및 평가 모의실험 결과, 요인별 지표변수의 수가 3개이고 요인계수가 0.6 이하인 모든 경우와 요인계수가 0.4로 낮으면서 동시에 요인별 지표변수의 수가 6개 이하인 경우 심각한 추정 오류를 나타내었다. 현재 분석 대상인 HRQoL 도구 중 반영적 측정구조를 포함하는 SF-36과 COHIP의 자료들은 요인계수가 비교적 크고 요인 당 지표변수의 수도 커서 예상되는 오류수준이 크지 않을 것으로 판단하였다.

셋째, PLS-SEM을 이용하여 HRQoL 측정도구에 대한 실증분석을 실시한

결과 모형이 잘 적합되었다. 상기한 HRQoL 도구에 대한 PLS-SEM 적용 결과 반영적 측정모형은 내적일관성, 지표변수 신뢰성, 수렴타당성 및 판별 타당성이 충족되었고 형성적 측정모형을 가지는 차원 역시 다중공선성 및 외부 가중치 타당성 면에서 적절하였다. 일차 구성개념들이 형성적 관계로 정의된 이차 계층 구조를 포함한 전체 구조방정식에도 PLS-SEM이 잘 적용되었다.

넷째, 여러 차원들을 다양한 방법으로 통합하여 HRQoL로 요약점으로 표현하여 비교한 결과, SF-36 및 COHIP 도구의 경우 PLS-SEM모형에서 나타난 차원 간을 통합하는 가중치를 이용한 경우 차원점수 평균이나 전체 문항점수 평균과의 상관 정도가 매우 높은 것으로 나타났다. PLS-SEM 적용 결과 산출된 가중치로 계산한 요약점수는 EQ-5D 도구의 선호도 기반 점수와 비교하였을 때에도 상당한 정도로 유사하였다. 따라서 본 연구의 결과에 따르면 사용자의 편리성 면에서 유리한 합산 점수나 평균 점수를 구성개념의 요약점으로 사용하는 것이 대체로 적절하다고 사료되었다.

본 연구에서는 대표적인 HRQoL 도구들을 분석함에 있어서 전체 측정도구의 구조를 아우르는 통계적 모형을 적용하였다. 통계적 모형 적용의 장점은 다음과 같다.

첫째, 측정모형의 타당성 평가가 가능하다. 모형 적용 후 분석 결과에 대하여 다양한 신뢰도 및 타당도 검증방법을 적용함으로써 전체 측정구조의 적합도, 개별 변수의 신뢰도 및 구성개념의 타당도 등을 평가할 수 있다. 둘째, 다양한 표본에 대한 분석 결과를 비교 평가할 수 있고 교차타당도 등의 검증을 통하여 측정모형의 일반화가 가능하다는 장점이 있다. 교차타당화를 시도하고 상이한 분석 결과에 대하여 차이의 원인을 규명하는 과정은 해당 도구의 측정모형에 대한 일반화를 위한 노력의 출발점이 될 수 있다.

본 연구에서는 HRQoL 도구의 측정모형에 PLS-SEM을 적용하였는데 측정모형 수립이라는 관점에서 볼 때 몇 가지 의문점이 제기되었다.

첫째, 모형에 함께 포함된 다른 변수로 인하여 추정치의 큰 변화가 나타난다는 점이다. 모형에 포함된 다른 변수의 변화에 따라 측정모형 자체가 일관된 결과를 산출하지 못한다면 정형화된 측정모형으로 사용하기에 어려움이 있을 수 있다.

둘째, 반영적 측정모형의 경우, 수렴타당도 기준이 AVE 0.5인데 이는 전체 분산의 50%가 구성개념과 무관한 고유분산에 해당하는 경우이다. 이런 경우에도 지표변수의 전체 변동을 포함하는 복합지수(composite)를 구성개념 값으로 이용하는 것은 상호 모순적인 면이 있다.

셋째, 형성적 모형의 수렴타당성을 평가하는데 중복 분석이 제안되었는데 본 연구에서 나타난 바와 같이 서로 다른 방법으로 조사한 유사한 구성개념간의 구조계수가 0.8을 상회하기는 현실적으로 쉽지 않을 것으로 보인다.

넷째, SF-36과 COHIP와 같이 이차 계층구조인 도구에서 형성적 관계로 생성되는 이차 구성개념을 가지는 경우, 모형의 결정계수는 대체로 1이다. 따라서 모형의 예측력이나 적합도를 평가하는데 결정계수를 활용하기 어렵게 된다.

다섯째, PLS-SEM의 경우 반영적 측정모형의 적합 시 지표변수의 특수 분산 간 상관 설정이 불가능하다. 이를 모형에 반영하고 있는 CB-SEM과 달리 PLS-SEM에서는 이러한 특수분산 간 상관의 설정이 가능하지 않다.

이 연구주제와 관련한 후속연구로 필요하다고 여겨지는 연구는 첫째, 보건학 분야의 다양한 구성개념으로 그 범위를 확대하여 그들의 구조를 이론적으로 또한 경험적으로 밝힐 필요성이 있다. 둘째, 형성적 구조의 데이터 및 이차 계층구조를 포함하여 다양한 조건과 구조를 설정하고 데이터를 생성하여 모의 실험을 수행함으로써 CB-SEM 및 PLS-SEM 적용 결과를 비교하고 그 수행을 평가하는 연구가 필요하다. 셋째, 반영적 및 형성적 측정모형의 구별이 이론적 내용적인 면으로 한정되어 다소 주관적인 판단이 개입될 우려가 있는데 이를 객관화시킬 수 있는 계량적 판단 기준의 개발이 요망된다. 넷째, PLS-SEM 전

체 모형의 적합도 평가 방법을 정교화 하는 후속연구가 필요하다고 사료된다.

결론적으로 본 연구의 의의는 전통적으로 반영적 구조에만 의거하여 측정 구조를 분석하던 여러 HRQoL 측정도구에 대하여 형성적 및 반영적 관계를 적절하게 규정하고 올바르게 정의된 측정모형에 PLS-SEM을 적용하여 경험적으로 구현하였으며 구현된 모형의 결과를 기반으로 다차원적 HRQoL 측정도구의 차원을 통합하는 방법을 제안하였다는 점에 있다.

## 참 고 문 헌

### <국내 문헌>

- 강은정, 신호성, 박혜자, 조민우, 김나연(2006). EQ-5D를 이용한 건강수준의 가치평가. 보건경제와 정책연구, 12(2), 19-43.
- 김장현, 심경환, 이철성 역(2014). In Hair, J. F., Hult, G. T., Ringle, C. M. PLS구조모델의 이해. 서울: 피앤씨미디어.
- 김혜영, 홍수민, 정세환, 안용순(2010). 어린이 일상활동구강영향(COIDP) 한국어판의 타당성 평가. 대한구강보건학회지, 34(4), 473-481.
- 성태제(1999). 교육평가방법의 변화와 결과타당도. 교육학연구, 제37권 제1호, 197-218.
- 성태제(2009). 문항반응이론의 이해와 적용. 서울: 교육과학사. 7-16.
- 성태제(2013). 타당도와 신뢰도. 서울: 학지사.
- 윤강재, 김계연(2010). OECD 국가의 행복지수 산정 및 비교. 보건복지포럼, 159, 86-98.
- 이순목(2000). 요인분석의 기초. 서울: 교육과학사, 9-31.
- 이순목(2002). 사회과학을 위한 측정의 원리. 서울: 학지사.
- 이순목, 김한조(2011). 구조방정식 모형의 일반화 또는 집단차 연구를 위한 다집단 분석의 관행과 문제점. 사회과학(성균관대), 43(1), 63-112.
- 차보경(2016). 성인 남성의 건강 관련 삶의 질에 영향을 미치는 용인에 관한 경로 분석. 지역사회간호학회지, 27(4), 399-409.
- 한창완, 이은주, 세키타 야스요시, 코즈키 마사히로(2009). 구조방정식을 이용한 한국어판 Short-Form 36 Health Survey (SF-36) 척도의 구성타당도 검증. 한국케어매니지먼트 연구, 4, 45-59.

<국외문헌>

- Ahn, Y. S., Kim, H. Y., Hong, S. M., Patton, L. L., Kim, J. H., Noh, H. J. (2012). Validation of a Korean version of the Child Oral Health Impact Profile (COHIP) among 8- to 15-year-old school children. *International Journal of Paediatric Dentistry*, 22(4), 292-301.
- Albers, S. (2010). PLS and success factor studies in marketing. In Esposito Vinzi V. et al (Eds.), *Handbook of Partial Least Squares, Concepts, methods and application* (pp. 409-425). Verlag Berlin Heidelberg: Springer Publishing Company.
- Amato, S., Esposito Vinzi, V., Tenenhaus, M. (2004). A global goodness-of-fit index for PLS structural equation modeling. *Oral Communication to PLS Club*, HEC School of Management, France, March 24.
- Allen M. J., Yen W. M. (1979). *Introduction to measurement theory*. Long Grove, IL, USA: Waveland Press Inc., 56-94.
- AREA, APA, & NCME (1966). *Standard of educational and psychological testing*. Washington D. C.: American Psychological Association.
- AREA, APA, & NCME (1985). *Standard of educational and psychological testing*. Washington D. C.: American Psychological Association.
- AREA, APA, & NCME (1999). *Standard of educational and psychological testing*. Washington D. C.: American Psychological Association.
- Arlinghaus, A., Lombardi, D. A., Willetts, J. L., Folkard, S., Christiani, D. C. (2012). A structural equation modeling approach to fatigue-related risk factors for occupational injury. *American Journal of Epidemiology*,

176(7), 597–607.

- Bentler, P. M., Stein J. A. (1992). Structural equation models in medical research. *Statistical Methods in Medical Research*, 1, 159–181.
- Beran, T. N., Violato, C. (2010). Structural equation modeling in medical research: a primer. *BMC Research Notes*, 3, 267.
- Bernabé, E., Sheiham, A., Tsakos, G. (2008). A comprehensive evaluation of the validity of Child–OIDP: further evidence from Peru. *Community Dentistry and Oral Epidemiology*, 36(4), 317–325.
- Bogazzi, R. P., & Fornell, C. (1982). Theoretical concepts, measurements, and meaning. In C. Fornell (Ed.), *A second generation of multivariate analysis: Vol. 2. Measurement and evaluation* (pp. 24–38). New York: Praeger.
- Bollen, K. A., Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methodology*, 16(3), 265–284.
- Bollen, K. A. (2015). In defense of causal–formative indicators: A minority report. *Psychological Methods*. (On–line published)
- Brazier, J., Roberts, J., Deverill M. (2002). The estimation of a preference–based measure of health from the SF–36. *Journal of Health Economics*, 21(2), 271–292.
- Broder, H. L., McGrath, C., Cisneros, G. J. (2007a). Questionnaire development: face validity and item impact testing of the Child Oral Health Impact Profile. *Community Dentistry and Oral Epidemiology*, 35(Suppl 1), 8–19.
- Broder, H. L., Wilson–Genderson, M. (2007b). Reliability and convergent and

- discriminant validity of the Child Oral Health Impact Profile (COHIP Child's version). *Community Dentistry and Oral Epidemiology*, 35(Suppl 1), 20–31.
- Bucic, T., Gudergan, S. P. (2004). The impact of organizational settings on creativity and learning in alliances. *Cairn Info*, 7(3), 257–273.
- Buhi, E. R., Goodson, P., Neilands, T. B. (2007). Structural equation modeling: A primer for health education behavior researchers. *American Journal of Health Behavior*, 31(1), 74–85.
- Bullock, H. E., Harlow, L. L., Mulaik, S. A. (2009). Causation issues in structural equation modeling research. *Structural Equation Modeling*, 1(3), 253–267.
- Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In Marcoulides, G. A. (Ed.), *Modern methods for business research* (pp. 295–358). Mahwah, NJ: Lawrence Erlbaum.
- Cho, Y. I., Lee, S., Patton, L. L., Kim, H. Y. (2016). Confirmatory factor analysis of the Child Oral Impact Profile (Korean version). *European Journal of Oral Sciences*, 124(2), 172–178.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ, Lawrence Erlbaum Associates.
- Constitution of the World Health Organization (2005). In: World Health Organization: *Basic documents. 45th ed.* Geneva: World Health Organization, 1.
- Cronbach L. J. (1951). Coefficient alpha and the internal structure of test. *Psychometrika*, 16, 297–334.
- Diamantopoulos, A., Winklhofer, H. M. (2001). Index construction with

- formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38,269–277.
- Dijkstra, T., Schermelleh–Engel, K. (2014). Consistent partial least squares for nonlinear structural equation models. *Psychometrika*, 79(4), 585–604.
- Diener E. (1984). Subjective well–being. *Psychological Bulletin*, 95, 542–575.
- Edwards J. R., Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5(2), 155–174.
- Embretson, S. E., Reise, S. P. (2000). *Item response theory for psychology*. Mahwah; NJ, USA:Lawrence Erlbaum Associates, Inc. 249–272.
- Erickson, P. (1998). Evaluation of a population–based measure of quality of life: the Health and Activity Limitation Index (HALex). *Quality of Life Research*, 7(2), 101–114.
- Farrell, A. M., Rudd, J. M. (2009). Factor analysis and discriminant validity: a brief review of some practical issues. In: *Tojib, D. (Ed.). ANZMAC 2009 conference proceedings*.
- Feeny, D., Furlong, W., Boyle, M., Torrance, G. W. (1995). Multi–attribute health status classification systems. Health Utilities Index. *Pharmacoeconomics*, 7(6), 490–502.
- Fornell, C., Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to Consumer–Exit–Voice theory. *Journal of Marketing Research*, 19, 440–452.
- Fornell, C., Larker, D. F. (1981). Evaluating structural equation modelings with unobservable variables and measurement error. *Journal of Marketing research*, 18(1), 39–50.

- Gandek, B., Sinclair, S. J., Kosinski, M., Ware, J. E. (2004). Psychometric evaluation of the SF-36 Health Survey in medicare managed care. *Health Care Financing Review*, 25(4), 5-25.
- Glaser, A. W., Furlong, W., Walker, D. A., Fielding, K., Davies, K., Feeny, D. H., Barr, R. D. (1999). Applicability of the Health Utilities Index to a population of childhood survivors of central nervous system tumors in the U. K. *European Journal of Cancer*, 35(2), 256-251.
- Gudergan, S. P., Ringle, C. M., Wende, S., Will, A. (2008). Confirmatory tetrad analysis in PLS path modeling. *Journal of Business Research*, 61(12), 1238-1249.
- Gusi, N., Olivares, P. R., Rajendran, R. (2010). The EQ-5D Health-Related quality of life questionnaire. In *Preedy, V. R., Watson, R. R. (ed.) Handbook of disease burdens and quality of life measures*, 87-99.
- Hair, J. E., Ringle, C. M., Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, 19, 139-151.
- Hair, J. F., Sarstedt, M., Ringle, C. M., Mena, J. A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Marketing Science*, 40(3), 414-433.
- Hair, J. E., Sarstedt, M., Pieper, T., Ringle, C. M. (2012a). The use of partial least squares structural equation modeling in strategic management research: A review of past practices and recommendations for future applications. *Long Range Planning*, 45, 320-340.
- Hair, J. F., Sarstedt, M., Hopkins L., Kuppelweiser, V. G. (2014) Partial least squares structural equation modeling (PLS-SEM) – An emerging tool in

- business research. *European Business Review*, 26(2), 106–121.
- Hays, R. D., Sherbourne, C. D., Mazel, R. M. (1995). *User's manual for the Medical Outcome Study (MOS) Core Measures of health-related quality of life*. Santa Monica: RAND, 4–7.
- Hays, R. D., Revicki, D., Coyne, K. S. (2005). Application of structural equation modeling to health research outcomes research. *Evaluation & The Health Professions*, 28(3), 295–309.
- Henseler, J. (2017). Bridging design and behavioral research with variance-based structural equation modeling. *Journal of Advertising*, 46(1),178–192.
- Hopkins, K. D., Stanley, J. C., Hopkins, B. R. (1990). *Educational and psychological measurement and evaluation*. Boston: Allyn & Bacon.
- Hoyt, C. (1941). The reliability obtained by analysis of variance. *Psychometrika*, 6, 153–160.
- Hunt, S. M., McEwen, J. (1980). The development of a subjective health indicator. *Sociology of Health and illness*, 2, 231–246.
- Hwang, H., Malhotra, N., Kim, Y., Tomiuk, M. A., Hong, S. (2010). A comparative study on parameter recovery of three approaches to structural equation modeling. *Journal of Marketing Research*, 47(4), 699–712.
- Jarvis, C. B., MacKenzie, S. B., Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecifications in marketing and consumer research. *Journal of consumer research*, 30, 199–218.
- Jenkinson, C., Layte, R., Lawrence, K. (1997). Development and testing of

- the Medical Outcomes Study 36-item Short Form health survey summary scale scores in the United Kingdom. *Medical Care*, 35(4), 410–416.
- Jo M. W., Yun, S. C., Lee, S. I. (2008). Estimating quality weights for EQ-5D health states with the time trade-off method in South Korea. *Value In Health*, 11(7), 1186–1189.
- Joreskog, K. G., Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70, 631–639.
- Joreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, 43(4), 443–447.
- Kuder, G. F., Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2, 151–160.
- Kaplan, R. M., Anderson, J. P. (1996). The general health policy model: An integrated approach. In Spiker, S. (ed.). *Quality of life and pharmacoeconomics in clinical trials (2<sup>nd</sup> ed.)*. Philadelphia: Lippincott–Raven Publishers, 309–322.
- Kaplan, R. M., Ganiats, T. G., Sieber, W. J., Anderson, J. P. (1998). The Quality and Well-Being scale: critical similarities and differences with SF-36. *International Journal for Quality in Health Care*, 10(6), 509–520.
- Kaplan, R. M., Ries, A. L. (2005). Quality of life as an outcome measure in pulmonary diseases. *Journal of Cardiopulmonary Rehabilitation*, 25, 321–331.
- Kind P. (1997). The performance characteristics of EQ-5D, a measure of health related quality of life for use in technology assessment [abstract]. *Annual Meeting of International Society of Technology Assessment in*

*Health Care*, 13(5), 81.

- Kontodimopoulos, N., Pappa, E., Niakas, D., Yfantopoulos, J., Dimitrakaki, C., Tountas, Y. (2008). Validity of EuroQoL (EQ-5D) instrument in a Greek general population. *Value Health*, 11(7), 1162–1169.
- Kupek, E. (2006). Beyond logistic regression: structural equation modeling for binary variables and its application to investigating unobserved confounders. *BMC Medical Research Methodology*, 6, 13.
- Linton, M. J., Dieppe P., Medina-Lara A. (2015). Review of 99 self-report measures for assessing well-being in adults: exploring dimensions of well-being and developments over time. *British Medical Journal Open*, 6e010641.
- Marsden, P. V. (1982). A note on block variables in multi-equation models. *Social Science Research*, 11, 127–140.
- McEwen, J. (1992). The Nottingham Health Profile. In: Walker, S. R., Rosser, R. M. (eds.). *Quality of life assessment: Key issues for the 1990s*. Dordrecht, Kluwer Academic Publishers.
- McHorney, C. A., Ware, J. E., Raczek, A. E. (1993). The MOS 36-item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care*, 31(3), 247–463.
- Mehrens, W. A. (1997). The consequences of consequential validity. *Educational Measurement: Issue and Practice*, 16(2), 16–18.
- Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin*. 100, 398–407.
- Mtaya, M., Åstrøm, A. N., Tsakos, G. (2007). Applicability of an

- abbreviated version of the Child–OIDP inventory among primary school children in Tanzania. *Health Quality of Life Outcomes*, 5, 40.
- Nully, J.C., Bernstein, I. (1994) *Psychometric theory*. New York (NY): McGraw–Hill p.217
- Partterson, T. L., Kaplan, R. M., Grant, I., Semple, S. J., Moscona, S., Koch, W. L., Harris, M. J., Jeste, D. V. (1996). Quality of well–being in late–life psychosis. *Psychiaty research*, 63(2–3), 169–181.
- Peng, D. X., Lai, F. (2012). Using partial least squares in operation management research: A practical guideline and summary of past research. *Journal of Operations Management*, 30(6), 467–480.
- Petrou, S., Morrell, J., Spiby, H. (2009) Assessment the empirical validity of alternative multi–attitude utility measures in maternity context. *Health and Quality of Life Outcomes*, 7, 40.
- Podsakoff, N. P., Shen, W., Podsakoff, P. M. (2006). The role of formative measurement models in strategic management research: Review, critique, and implications for future research. *Research Methodology in Strategy and Management*, 3, 197–252.
- Popham, W. J. (1997). Consequential validity: right concern–wrong concept. *Educational Measurement: Issue and Practice*, 16(2), 9–13.
- Rabe–Hesketh, S., Skrondal, A. (2008). Classical latent variable models for medical research. *Statistical Methods in Medical Research*, 17, 5–32.
- RAND Home page. Accessible at  
[https://www.rand.org/health/surveys\\_tools/mos/36-item-short-form/scoring.html](https://www.rand.org/health/surveys_tools/mos/36-item-short-form/scoring.html).
- Rapaport, M. H., Bazzetta, M. A., McAdams, L. A., Patterson, T., Jeste, D. V. (1996). Validation of the scale of functioning in older outpatients with

- schizophrenia. *American Journal of Geriatric Psychiatry*, 4(3), 218–228.
- Rigdon, E. E. (2012). Rethinking partial least squares path modeling: in praise of simple methods. *Long Range Planning*, 45, 341–358.
- Ringe, C. M., Sarstedt, M., Straub, D. W. (2012). A critical look at the use of PLS–SEM in MIS Quarterly. *MIS Quarterly*, 36(1), 3–14.
- Rosner, B. (2006). *Fundamentals of biostatistics (6<sup>th</sup> Eds.)*. Belmont: Thomson Brooks/Cole.
- Ross, S. M. (2012). *Introduction to probability and statistics for engineers and scientists (4th Eds.)*. Seoul: Elseveiser Korea L.L.C, 268–274.
- Rossiter J. R. (2002). The C–OAR–SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19, 305–335.
- Roy, S., Tarafdar M., Ragu–Nathan, T. S., Marsillac E. (2012). The effect of misspecification of reflective and formative constructs in operation and manufacturing management research. *Electronic Journal of Business Research Methods*, 10(1), 34–51.
- Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: When the bias lies! *Journal of Business Research*, 69, 3998–4010.
- Scott–Lennox, J. A., Wu, A. W., Boyer, J. G., Ware, J. E. (1999). Reliability and validity of French, German, Italian, and UK English translations of the Medical Outcomes Study HIV Health Survey. *Medical Care*, 37(9), 908–925.
- Schipper, H. (1990) Guidelines and caveats for quality of life measurements in clinical practice and research. *Oncology*, 4(5), 51–57.
- Schipper, H., Clinch, J. J., Olweny, C. L. M. (1996). Chapter 2. Quality of

- life studies: definition and conceptual issues. In Spilker, B. *Quality of life and phamacoeconomics in clinical trials (2<sup>nd</sup> Ed)*. Philadelphia Pennsylvania: Lippincott–Raven Publishers, 11–24.
- Schneeweiss, H., Mathes, H. (1995). Factor analysis and principal components. *Journal of Multivariate Analysis*, 55(1), 105–124.
- Shafie, A. A., Hassali, M. A., Liau, S. Y. (2011). A cross–sectional validation study of EQ–5D among the Malaysian adult population. *Quality Life Research*, 20(4), 593–600.
- Shepard, L. A. (1997). The centrality of test use of consequences for test validity. *Educational Measurement: Issue and Practice*, 16(2), 5–8.
- Simonetto A. (2012). Formative and reflective models to determine latent construct, *presented at 46th Meeting of Italian Statistical Society, Rome, 2012*.
- Spearman C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72–101.
- Stephenson, M. T., Holbert, R. L., Zimmermen, R. S. (2006). On the use of structural equation modeling in health communication research. *Health Communication*, 20(2), 159–167.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 667–680.
- Tekwe, C. D., Carter, R.L., Cullings H. M., Carroll R. J. (2014). Multiple indicators, multiple causes measurement error models. *Statistics in Medecin*, 33(25), 4469–4481.
- Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.–M., Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis*, 48(1), 159–205.

- The EuroQol group (1990). EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy*, 16, 199–208.
- Tobias, R. D. (1995). An introduction to partial least square regression. *SUGI proceedings*. Accessible at <https://support.sas.com/rnd/app/stat/papers/abstracts/pls.html>.
- Vanderweele, T. J. (2012). Invited Commentary: Structural equation models and epidemiologic analysis. *American Journal of Epidemiology*, 176(7), 608–612.
- Velicer, W. F. (1977). An empirical comparison of the similarity of principal component, image, and factor patterns. *Multivariate Behavioral Research*, 12, 3–22.
- Velicer, W. F., Peacock, A. C., & Jackson, D. N. (1982). A comparison of component and factor patterns: A Monte Carlo approach. *Multivariate Behavioral Research*, 17, 371–388.
- Vinzi, V. E., Trinchera, L., Amato, S. (2010). PLS path modeling: From foundation to recent developments and open issues for model assessment and improvement. In Vinzi, V. E., Chin, W. W., Henseler, J., Wanf, H. (Eds). *Handbook of partial least squares—Concepts, methods and applications* (pp. 47–82). Heidelberg: Springer.
- Violato, C., Hecker, K. G. (2007). How to use structural equation modeling in medical education research: A brief guide. *Teaching and Learning in Medicine*, 19(4), 362–371.
- Ware, J. E., Sherbourne, C. D. (1992). The MOS 36-item Short-Form Health Survey (SF-36): I. Conceptual framework and item selection. *Medical Care*, 30(6), 473–483.

- Wetzels, M., Odekerken–Schroder, G., Oppen, C. V. (2009). Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. *MIS Quarterly*, 33(1), 177–195.
- Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters. *Multivariate Behavioral Research*, 28, 263–311.
- Wilson, I. B., Cleary, P. D. (1995). Linking clinical variables with health–related quality of life: A conceptual model of patient outcomes. *Journal of the American Medical Association*, 273(1), 59–65.
- Wold, H. O. A. (1982). Soft modelling: The basic design and some extensions. In K. G. Joreskog & H. O. A. Wold (Eds.). *Systems indirect observations: Part II* (pp. 1–54). Amsterdam: North–Holland.
- World Health Organization (1948). *Constitution of the World Health Organization*. Geneva: WHO basic Document.
- Yusuf, H., Gherunpong, S., Sheiham, A., Tsakos, G. (2006). Validation of English version of the Child–OIDP index, an oral health–related quality of life measure for children. *Health Quality Life Outcomes*, 4, 38.

# ABSTRACT

An Application of Partial Least Square Structural Equation Models (PLS-SEM) on Health-Related Quality of Life(HRQoL) Measures

Hae-Young Kim

Department of Psychology

Graduate School of Sungshin Women's University

Interest on Health-Related Quality of Life (HRQoL) measures is increasing as the comprehensive evaluation of health status including objective clinical examination and subjective report of patients predominates gradually. HRQoL measures are multidimensional inventories which are composed of multiple dimensions with reflective measurement structures, and a secondary factor summarizing lower dimensions as formative relationships. However, traditionally HRQoL measures have been treated as having only reflective structures and reflective relationships according to the psychometric methodology.

This study investigated the theoretical contextual structures of several major HRQoL measures such as Short Form-36(SF-36), EuroQoL-5D(EQ-5D), Oral Health Impact Profile for Children(COHIP), and Oral Impacts on Daily Performances for Children (COIDP) in a viewpoint of reflective and formative measurement models. Also we empirically applied

PLS–SEM models on data of several major HRQoL measures.

First, we meticulously considered the contents of several major HRQoL measures and postulated the relevant relationships between dimensions and the upper construct. According to the results, we established appropriate conceptual models with reflective or formative relationships and used the model in the following empirical analyses.

Second, we assessed the degree of estimation errors of PLS–SEM models when applied on data from reflective structures by simulation of data. According to the analysis results, there were significant amount of estimation errors in all the conditions with small number of indicators of 3, and with small number of indicators of 6 and low factor coefficients of 0.4 coincidentally. However, our empirical data appeared not to have such disadvantageous conditions.

Third, it was appropriate to apply PLS–SEM models on empirical HRQoL data.

Forth, the summary HRQoL measures, using various methods including application of weights estimated by PLS–SEM, showed similar results and had high inter–correlations.

There were several questions related to the application of PLS–SEM.

First, there was considerable variation in the size of factor coefficients or outer weights as models included different variables. Second, a reflective model with 0.5 of average variation extracted(AVE) can be accepted. However it may be unreasonable using the composite score summarizing all the variation of the indicators. Third, one condition to evaluate formative measurement models is to obtain at least 0.64 of R–square value in

redundancy analysis. However, it may not be attainable in actual situations.

In conclusion, the importance of this study lies on the fact that we tried to apply appropriate relation of reflective and formative measurement models on HRQoL measures and empirically examined the results by using PLS-SEM. In addition, we suggested the summary measure of HRQoL using weights from PLS-SEM models.

*Key words: Health-related quality of life (HRQoL), Oral health-related quality of life (OHRQoL), Partial Least Square Structural Equation Modeling (PLS-SEM), Covariance-based Structural Equation Modeling (CB-SEM), Formative measurement model, Composite indicator variable.*

## 부록 1. 고전검사이론 (Classical Test Theory)

### 1. 진점수이론

전통적으로 대부분의 표준적 검사들은 고전검사이론에 의거하여 검사도구를 개발하고 평가하였다. 고전검사이론의 핵심인 진점수이론(true-score theory)은 단순한 모델로서 측정오차가 관찰점수에 어떻게 영향을 미치는가를 효과적으로 설명하였다. 진점수이론은 몇 개의 주요한 가정에 기반하고 있으며 만약 가정이 합리적이라면 그 모형을 적용한 결과에 의한 결론도 역시 합리적이라고 간주한다.

$X$  를 관찰점수,  $T$  를 진점수 및  $e$ 를 오차로 하여 고전 진점수이론의 가정들을 정리하여 보면 다음과 같다:

- 1)  $X=T+e$  : 관찰점수는 진점수와 오차로 구성된다.
- 2)  $E(X)=T$ : 무수히 반복 측정된 관찰점수의 평균은 진점수와 같다. 이는 오차가 평균이 0인 무선오차를 가정하기 때문이다
- 3)  $\rho_{Te} = 0$  : 진점수와 오차점수는 무관하다. 즉 능력이 서로 달라 진점수가 높거나 낮더라도 오차가 작거나 크지 않다
- 4)  $\rho_{ee'} = 0$  : 한 검사에서 얻은 오차점수는 다른 검사에서 얻은 오차점수와 상관이 없어 상호독립적이다.
- 5)  $\rho_{T_1e_2} = 0$  : 한 검사에서 얻은 진점수와 다른 검사에서 얻은 오차와의 상관은 0이다.
- 6)  $T = T', \sigma_e^2 = \sigma_{e'}^2$  : 한 검사에서 얻은 진점수가 다른 검사에서 얻은 진점수와 같고 두 검사의 관찰점수의 분산이 동일하다면 동형검사(parallel test)로 간주한다.

## 2. 검사도구의 타당도(validity) 평가

타당도(validity)는 검사도구를 평가함에 있어 “검사도구가 측정하고자 하는 목표 대상을 얼마나 충실히 측정하였는지”로 정의된다(성태제, 2013). 따라서 타당도는 항상 무엇을 측정하는가와 타당한 정도를 언급하게 되며, 모든 내용에 타당한 검사는 존재하지 않고 타당도가 있다거나 없다는 것이 아니라 어느 정도로 타당한가 하는 것을 평가한다.

타당도의 개념은 이론적으로 계속 발전하여 최근에는 검사에서 얻은 결과를 가지고 검사의 타당성의 근거를 제시하는 과정(AREA, APA, NCME, 1985) 또는 검사 목적에 의하여 설명된 검사점수의 해석에 대하여 근거나 이론이 지지하여 주는 정도(AREA, APA, NCME, 1990) 로 정의되고 있다.

지금까지 많은 학자들에 의하여 다양한 명칭의 타당도가 제안되었다. Anastasi(1954)와 Chronbach(1961) 등은 여러 번에 걸쳐 타당도의 종류를 세 종류 또는 네 종류 등으로 약간씩 변경하여 제시하였는데 대체적으로는 AREA(American Educational Research Association), APA(American Psychological Association), 및 NCME(National Council on Measurement in Education)에 의한 분류 방법을 따르고 있다(표2). AREA, APA, NCME(1999)에서는 계량적 접근에 의한 타당도 수준을 초월하여 질적인 접근에 의한 타당도의 개념으로 넓혀 ‘~에 기초한 근거(evidence based on~)’라는 단어를 사용하여 검사내용에 기초한 근거’, ‘반응과정에 기초한 근거’, ‘내적 구조에 기초한 근거’, ‘다른 변수와의 관계에 기초한 근거’, ‘검사결과에 기초한 근거’ 등 다섯 종류의 타당도로 분류하였다. ‘다른 변수와의 관계에 기초한 근거’에는 ‘수렴 및 판별 근거’, ‘검사-준거 관련성’, ‘타당도 일반화’가 있다.

Table 1. American Educational Research Association (AREA), American Psychological Association (APA), 및 National Council on Measurement in Education (NCME) 에서 제시한 타당도(성태제, 2013)

AREA, APA, NCME (1966)	AREA, APA, NCME (1985)	AREA, APA, NCME (1999)
•Content validity	•Content-related evidence of validity	•Evidence based on test content •Evidence based on response process •Evidence based on internal content
•Construct validity	•Construct-related evidence of validity	•Evidence based on relations to other variables - Convergent and discriminant evidence
•Criterion-related validity	•Criterion-related evidence of validity	•Test-criterion relationships •Validity generalization •Evidence based on consequences of testing

## 2.1 내용타당도 (content-related validity): 검사 내용에 기초한 근거

내용타당도(검사내용에 기초한 근거, evidence based on test content)란 논리적 사고에 입각한 논리적인 분석과정으로 판단하는 주관적인 타당도로 검사 내용 전문가에 의하여 주관적으로 검사가 측정하고자 하는 속성을 제대로 측정하였는지를 판단하여 검사도구의 타당도를 평가하는 방법이다(성태제, 2013). 검사의 문항, 질문, 목적이 측정을 위하여 규정된 내용 영역이나 전체를 얼마나 잘 대표하느냐의 정도를 평가하는 과정이다.

## 2.2 반응과정에 기초한 근거

반응과정에 기초한 근거(evidence based on response process)란 피험자의 응답을 분석하여 반응과정에 대한 분석을 통하여 검사가 측정하고자 하는 구인과 피험자의 반응이 일치하는가에 근거하여 타당도를 검증하는 것이다.

## 2.3 구인타당도(construct validity): 내적 구조에 기초한 근거

구인타당도(내적 구조에 기초한 근거, evidence based on internal structure)란 조작적으로 정의되지 않은 인간의 심리적 특성이나 성질을 심리적 구인으로 분석하여 조작적 정의를 부여한 후, 검사의 점수가 부여한 심리적 구인으로 구성되어 있는가를 검증하는 방법이다. 구인타당도는 구성타당도라고도 한다.

구인(construct)이란 심리적 특성을 설명하기 위하여 있을 것이라고 가정하는 심리적 요인을 말한다. 예를 들어 창의력을 측정할 때, 창의력은 민감성, 이해성, 도전성, 개방성, 자발성, 그리고 자신감으로 구성되어 있다고 한다면, 이와 같은 심리적 요인을 구인이라고 하고 그 검사도구가 이 구인들을 실제로 측정하고 있는지를 밝히는 것이 구인타당도를 검증하는 것이다.

구인타당도를 측정하는 절차는 첫째, 알지 못하는 심리적 특성을 구성하는 구인이 무엇인지 밝히고 둘째, 구인들과 관련된 구인들을 측정하는 문항을 제작하여 검사를 만든 후 셋째, 검사를 실시하여 응답자료인 검사점수를 얻고 넷째, 검사점수가 심리적 구인들을 포함하고 있는지 즉, 검사가 측정하고자 하는 구인들을 모두 측정하였는지를 양적인 방법으로 검증한다. 내적 구조에 기초한 근거를 확인하기 위한 방법으로는 종전의 구인타당도를 검증하는 방법으로 사용되는 요인분석을 사용할 수 있다(성태제, 2013)

## 2.4 다른 변수에 기초한 근거

### (1) 수렴 및 판별 근거

수렴 및 판별 근거(convergent and discriminant evidence)'에서는 해당 검사 점수와 유사한 구인을 측정하는 측정치를 이용하여 수렴근거를 제공하고 반대로 해당 검사점수와 다른 구인을 측정하는 측정치와 비교하여 판별근거를 제공한다. 여러 방법으로 측정한 동일한 특성의 점수는 그들 간의 상관이 높을 때 수렴근거가 있다고 보며 서로 다른 특성을 측정한 점수는 그들 간의 상관이 높지 않을 때 판별근거가 있다고 보는 것이다.

### (2) 검사-준거 관련성

'검사-준거 관련성(test-criterion relationship)'은 검사 점수가 얼마나 정확하게 준거 수행을 예측하는가에 초점을 맞추어 평가한다. 준거변수는 검사 학교 운영자나 회사경영자가 검사 사용자들이 관심을 가지는 속성이나 결과로서 예를 들면 미래의 학업수행능력이나 미래의 업무수행정도 등이다.

#### (가) 예측타당도

예측타당도(준거와 관련된 예측근거, criterion-related evidence of validity; 예측근거, predictive evidence)는 제작된 검사와 미래의 어떤 행위와의 관계로 평가하는 타당도 이다.

### (나) 공인타당도

공인타당도(공인근거, concurrent evidence)는 얻은 검사점수와 준거로 기존에 타당성을 인정받고 있는 검사로부터 얻은 점수와의 관계에 의하여 입증하는 타당도이다.

### (3) 타당도 일반화

타당도 일반화(validity generalization)은 검사-주거 관련성에 기초한 타당성의 근거가 다양한 상황에서 같거나 유사한 정도로 나타나 일반적으로 타당성을 인정할 수 있는 정도를 말한다. 다양한 상황은 측정 방법상의 차이, 직업이나 교육과정의 유형, 피험자 유형, 타당화 연구 시기 등을 상정할 수 있다.

## 2.5 결과타당도: 검사 결과에 기초한 근거

결과타당도(검사결과에 기초한 근거, evidence based on consequences of testing)는 검사결과에 대한 가치 판단으로 검사 결과가 평가목적과 부합하는 실제적 사회적 결과를 가져오는지 평가하는 것이다. 예를 들면 학생들의 수행평가가 학생들의 학습동기를 유발하고 의도한 대로 긍정적인 효과가 나타나는지 등을 고려하여 검사의 결과타당도를 평가하는 방법이다.

결과타당도가 타당도의 종류에 포함되는 것이 옳으냐 하는 것에 대해서 검사결과에 대한 책임과 배려 등의 중요성이 크므로 포함시켜야 한다는 주장(Shepard, 1997;성태제, 1999)과 개념의 모호성에 의한 혼란이나 외부적 가치판단이 게재하는 문제점 등의 이유로 타당도에 포함시키지 않아야 한다는 주장

(Popham, 1997; Mehrens, 1997)이 공존하고 있다.

### 3. 검사도구의 신뢰도 평가

신뢰도는 검사도구가 어떤 속성을 변함없이 일관성 있게 측정하는 정도를 나타낸다. 검사도구의 타당성이 입증되었다면 신뢰성이 고려되어야 한다. 어떤 검사도구가 속성을 측정할 때마다 같은 점수를 얻는다면 이 검사도구가 신뢰성이 있는 검사가 된다. 신뢰도(reliability)란 측정하려 하는 것을 얼마나 안정적으로 일관성 있게 측정 하였느냐를 의미하며 만약 측정 시 오차가 크다면 일반적으로 신뢰도는 낮아진다.

현재까지 여러 신뢰도의 추정방법이 개발된 바 있다. 일찍이 Spearman(1904)이 상관계수에 의하여 신뢰도계수(reliability coefficient)의 개념을 설명한 바 있으며 그 이후에 개발된 신뢰도의 종류는 재검사 신뢰도, 동형검사 신뢰도 및 내적일관성 신뢰도 등을 들 수 있다. 내적일관성 신뢰도에는 반분검사 신뢰도(Spearman and Brown, 1910), Kuder-Richardson 20(Kuder-Richardson, 1937), Kuder-Richardson 21, Hoyt의 신뢰도(Hoyt, 1941), 그리고  $\alpha$  계수 (Cronbach, 1951)가 포함된다.

#### 3.1 재검사 신뢰도(test-retest reliability)

재검사신뢰도는 동일한 검사를 동일한 피험자 집단에 일정 시간 간격을 두고 두 번 실시하여 두 검사 점수의 상관관계에 의하여 신뢰도를 검증하는 방법이다. 상관계수에 의하여 신뢰도 계수가 산출되며 이는 검사도구의 안정성을 측정하게 된다. 재검사신뢰도는 검사를 두 번 실시하며, 시험 간격에 의하여 신뢰

도 계수가 변화하는 단점을 지니고 있다.

일반적으로 피험자의 기억이 소멸되는 기간으로 2주에서 4주로 설정하나 검사도구의 특성, 측정 내용에 따라 달라 질 수 있다. 검사를 두 번 시행하는데 따르는 문제점과 학습효과 및 대상자의 속성 변화와 관련하여 검사시간 간격 설정에 따른 재검사신뢰도의 변화라는 큰 단점 때문에 사용이 감소되고 있는 추세이다.

### 3.2 동형검사 신뢰도(parallel-form reliability)

동형검사신뢰도는 검사도구의 신뢰도를 검증하기 위하여 동형의 검사도구를 제작한 뒤 동일 피험자 집단에게 검사를 실시하여 두 검사점수의 상관계수에 의하여 신뢰도를 검증하는 방법이다. 동형검사신뢰도 역시 두 검사 점수의 상관계수로 추정하므로 재검사신뢰도를 추정하는 절차와 동일하다. 다만 한 검사의 점수와 다른 동형검사의 점수라는 내용만 재검사신뢰도와 다를 뿐 추정절차는 똑같으므로 동형검사신뢰도 추정 절차를 생략한다.

동형검사신뢰도 추정은 재검사신뢰도의 시험 간격 설정의 문제점을 해결하나, 두 번 검사를 시행해야 하는 어려움이 있으며 더욱 큰 단점은 동일 형태, 동일 문항 수, 동일 문항난이도 및 변별도 등의 조건이 만족되는 동형검사의 제작이 현실적으로 용이하지 않다는 점이다. 동형검사신뢰도는 두 검사의 유사성을 측정한다.

### 3.3 내적일관성 신뢰도(internal consistency reliability)

내적일관성신뢰도는 검사를 구성하고 있는 문항간의 일관성을 측정하므로 검사를 두 번 실시하지 않고 검사의 신뢰도를 추정할 수 있는 장점을 지니고 있다. 내적일관성신뢰도는 검사를 구성하는 두 부분 검사 사이의 유사성에 의해 추정되는 반분검사신뢰도와 문항간의 측정의 일관성에 의해 추정되는 문항내적 일관성신뢰도로 나눌 수 있다.

반분검사신뢰도(split-half reliability)는 한번 실시한 검사 점수를 두 부분으로 나누어 두 부분간 점수의 상관계수를 계산한 후, Spearman과 Brown이 제안한 공식에 의하여 신뢰도를 검증하는 방법이다. 검사점수를 양분하는 방법으로 전후법, 홀수-짝수법, 문항특성에 따른 양분법, 무선 양분법이 있다. 양분된 검사는 원래의 검사에 비하여 신뢰도를 과소 추정하므로 Spearman Brown이 이를 교정하기 위하여 다음의 공식을 제안하였다.

$$r_{XX'} = 2r_{YY'} / (1+r_{YY'}),$$

$r_{XX'}$ : 원 검사의 신뢰도,  $r_{YY'}$ : 반분된 검사의 신뢰도.

반분검사신뢰도는 한번의 검사 실시로 신뢰도를 추정할 수 있는 장점을 지니고 있으나, 검사를 양분하는 방법에 따라 신뢰도 계수가 변하는 단점을 지니고 있다.

문항내적일관성신뢰도(internal consistency reliability)는 검사를 한 번 실시하되 검사를 양분하지 않고 문항간의 일치정도를 추정하여 검사의 신뢰성을 검증하는 방법으로 KR-20 (Kuder-Richardson 20), KR-21 (Kuder-Richardson 21), Hoyt 신뢰도, 및 alpha 계수 등이 포함된다.

KR-20과 KR-21은 Kuder와 Richardson(1937)에 의하여 제안된 방법이며, KR-20은 문항 점수가 0과 1일 때 사용하며, KR-21은 문항 점수가 Likert

척도와 같이 1,2,3,4,5 점과 같을 때 검사의 신뢰도를 추정하는 방법이다. Hoyt 신뢰도는 1941년 분산분석의 반복설계를 이용하여 Hoyt에 의하여 제안된 방법이며, alpha( $\alpha$ )계수는 Cronbach(1951)에 의하여 제안된 방법으로 신뢰도를 표기할 때 가장 자주 쓰이는 방법이다. Hoyt 신뢰도와 alpha 계수는 문항 점수가 0과 1이거나 1,2,3,4일 때에 모두 사용된다.

일반적으로 인지영역 검사로서 학업성취도 검사에서 정답이 있는 선택형 문항일 경우 문항 점수가 0과 1로 부여된다. 같은 응답자료를 가지고 검사도구의 신뢰도를 KR-20, Hoyt, alpha( $\alpha$ )값에 의하여 계산하면 신뢰도 계수가 같다. 이는 세 신뢰도 추정공식은 다소 다르나 신뢰도를 규명하는 이론적 배경이 동일하기 때문이다. 여러 내적일관성 신뢰도중 가장 널리 사용되는 alpha( $\alpha$ )값을 추정하는 공식은 다음과 같다

$$\alpha = \frac{n}{n-1} \left( 1 - \frac{\sum S_i^2}{\sum S_x^2} \right)$$

n : 문항 수  $s_i^2$  : 문항 점수 분산  $s_x^2$  : 총점의 분산

문항점수 분산은 피험자들이 각 문항에 응답하여 얻은 점수들의 분산이며, 총점의 분산은 피험자들이 검사에서 얻은 총점의 분산을 의미한다. 분산이란 각 개인의 점수가 평균점수로부터 떨어진 정도를 나타내는 것으로 다음의 공식에 의하여 계산한다.

$$S_x^2 = \frac{\sum (x - \bar{x})^2}{N}$$

X : 개인 점수 : 평균 점수 N : 피험자 수

이상에서 설명한 재검사신뢰도, 동형검사신뢰도, 내적일관성신뢰도 중 가장

과학적인 신뢰도 측정도구는 내적일관성신뢰도로서 그 중에서도 KR-20, Hoyt,  $\alpha$ ( $\alpha$ )계수 등이 대표적으로 사용된다.

신뢰도 계수를 평가하는데 있어 몇 가지 고려하여야 하는 요인이 있다.

첫째, 문항 수이다. 만약 적은 수의 문항으로 인간이 지니고 있는 속성을 측정할 때 보다 많은 수의 문항으로 검사를 실시할 때 측정의 오차를 줄일 수 있다.

둘째, 피험자가 정답을 맞힐 확률인 문항난이도가 적절하여야 한다. 검사가 너무 어렵거나 쉬우면 검사 불안이나 부주의가 발생하여 일관성 있는 응답을 하지 못하므로 신뢰도가 저하된다.

셋째, 피험자를 능력에 따라 구분할 수 있는 정도인 문항변별도가 높아야 한다.

넷째, 검사도구의 측정 내용이 보다 좁은 범위의 내용이어야 한다. 검사내용의 범위를 좁힐 때, 문항간의 동질성을 유지하기가 용이하다.

다섯째, 검사 시간이 충분하여야 한다. 이는 문항 수와 관계되는 문제이기도 하다. 충분한 시간이 부여될 때 응답의 안정성을 보장받을 수 있다(성태제, 2013).

## 부록 2. 요인분석 이론(Factor Analysis Theory)

요인분석은 사회과학과 생물과학에서 수집되는 다수 또는 대량의 자료를 처리해서 이론적으로 또는 내용적으로 의미 있는 소수의 변수를 추출하는 통계방법이다. 요약된 소수의 변수는 단순한 축약이 아니라 내용적으로 의미 있는 개념들이 되므로 이를 요인(factor)이라 부른다. 이 개념들은 이론적 구조에서 핵심적인 개념으로 사용되므로 이론변수라고 하며 직접 관찰하는 것이 아니라 측정변수 속에 들어 있는 알맹이로 간주되므로 잠재변수 또는 관찰되지 않는 변수라고 한다.

사회과학에서 사용하는 여러 구성개념들은 많은 경우, 직접적으로 측정할 수 없지만 여러 측정변수들에 공통적으로 작용하는 알맹이를 요약하여 평가하는 방법으로 측정한다. 예를 들면 지능은 직접 측정이 불가능하나 대상자에게 구체적인 문제를 다양하게 풀이하게 하고 그 풀이과정 속에 공통적으로 작용하는 잠재된 지능의 정도를 평가하여 간접적으로 측정 가능하다. 이런 경우 (공통)요인분석법(Common Factor Analysis, CFA)을 적용할 수 있다.

반면 사회과학이나 보건학 등에서 소수의 축약된 변수가 사물의 여러 측면을 대표하는 측정변수 또는 지표변수의 가중선형결합의 형태로 하나의 구성개념을 형성하기도 한다. 예를 들면 사회경제적 지위의 경우 사회경제적 지위를 결정짓는 서로 다른 측면을 나타내는 소득, 교육, 자산 등을 적절하게 합성하여 나타낼 수 있다. 이런 경우 주성분분석법(Principal Component Analysis, PCA)을 적용할 수 있다. 요인분석은 해당 구성개념과 관찰변수의 관계의 특성에 따라 CFA와 PCA로 구분된다.

## 1. 공통요인분석(Common Factor Analysis)

공통요인 분석은 관찰변수 간의 상관관계를 소수의 잠재변수에 의해서 설명하는 통계방법이다. 여러 관찰변수가 소수의 공통요인에 의하여 설명될 수 있는 관계라면 관찰변수 간의 상관이 클수록 그 변수들을 공통된 개념으로 요약하기 쉬워진다.

예를 들면 수학실력이 좋거나 나쁘다면 산술시험 점수는 수학실력과 연관되므로 다양한 종류의 산술시험 점수가 유사하게 좋거나 나쁠 수 있다. 반대로 개별 산술점수들 간의 상관이 높다면 수학실력이라는 하나의 공통된 구성개념으로 요약하기가 쉬울 것이다. 따라서 개별 산술점수들 간의 공통분을 나타내는 상관이 크다는 전제가 있어야 요약된 공통된 구성개념이 잘 측정될 수 있다고 본다.

이때 공통요인모형에서는 개별 산술점수는 공통요인인 수학실력과의 가중된 선형적 관계에 의해 표현될 수 있는데, 동시에 개별 산술 시험의 고유한 특성이 존재하므로 고유한 요인 역시 관계식에 포함되어 개별 산술점수는 다음과 같이 표현될 것이다.

$$X_1 = \lambda_1 F + U_1, X_2 = \lambda_2 F + U_2$$

$X_1$  와  $X_2$  는 산술점수이고  $F$ 는 공통요인인 수학실력이며  $U_1$  과  $U_2$  고유요인을 나타내고  $\lambda_1$  과  $\lambda_2$  는 요인계수이다.

공통요인  $F$ 가 표준화된 값으로 나타난다면, 두 변수  $X_1$  와  $X_2$  간 상관은 다음과 같이 계산할 수 있다.

$$\text{Corr}(X_1, X_2) = \text{Corr}(\lambda_1 F + U_1, \lambda_2 F + U_2) = \text{Corr}(\lambda_1 F, \lambda_2 F) = \lambda_1 \lambda_2$$

또한 한 변수의 분산은 다음과 같이 공통요인을 반영하는 분산과 고유요인을 반영하는 분산으로 나누어지게 된다.

$$Var(X_1) = Var(\lambda_1 F + U_1) = \lambda_1^2 Var(F) + Var(U)$$

이를 정규화된 산술점수를 사용하여 일반화하여 보면 다음 식들로 열거된다.

$$z_1 = \lambda_{11} F_1 + \lambda_{12} F_2 + \epsilon_1$$

....

$$z_9 = \lambda_{91} F_1 + \lambda_{92} F_2 + \epsilon_9$$

이를 매트릭스 형태로 요약하면 다음과 같다.

$$Z = \Lambda F + \epsilon$$

공통요인분석에서는 공통요인이 차지하는 분산이 일정 정도 이상으로 커야 개별 측정변수들이 동일한 공통요인을 측정하고 있음을 인정할 수 있으며 따라서 고유요인이 차지하는 분산은 일정 정도 이하로 작아야 한다.

## 2. 주성분분석(Principle Component Analysis)

주성분분석은 구성개념을 이끌어내고자 하는 목적은 공통요인분석과 유사하나 관찰변수들 간의 상관관계를 표현하는 것이 아니라 소수의 관찰변수들의 합성변수로 구성개념을 표현하는 방법이다. 따라서 공통적인 부분이 중요하지 않고 각각의 관찰변수가 의미하는 부분을 최대한 많이 포함하도록 합성변수를 구성하는 것이 주성분분석의 목표가 된다.

이 경우 구성개념은 각 측면을 잘 반영하는 측정변수들로 합성되므로 공통

요인이나 고유요인의 개념을 구별하여 인정하지 않고 단지 각 측정변수의 분산을 최대한 많이 표현하여 줄 수 있도록 소수의 구성개념을 만들어 낸다. 주성분분석 모형에서는 다음과 같이 구성개념이 관찰변수의 선형결합으로 정의된다.

$$P_1 = \gamma_{11}z_1 + \lambda_{12}z_2 + \dots + \lambda_{1n}z_n + \zeta_1$$

....

$$P_m = \gamma_{m1}z_1 + \lambda_{m2}z_2 + \dots + \lambda_{mn}z_n + \zeta_m$$

이를 통합하면 다음과 같다.

$$\sum_{j=1}^m P_j = \sum_{i=1}^n \sum_{j=1}^m \gamma_{ji}z_i + \sum_{j=1}^m \zeta_i$$

만약 공통요인모형에서 고유요인이 매우 작아서 무시할 수 있는 정도라고 한다면 주성분모형과 공통요인 모형이 수학적으로 동치가 된다. 그러나 일반적으로 고유요인의 존재가 무시할 정도로 작다고 볼 수 없으므로 현실적으로 공통요인모형과 주성분모형의 수학적 동치관계는 성립하지 않는다.

요약하면 공통요인모형은 측정변수 간 상관을 설명하는 공통요인을 추출하는 모형이고, 주성분모형은 측정변수들의 분산을 설명하기 위하여 측정변수의 선형결합으로서 주성분을 추출하는 모형이다. 따라서 공통요인모형에서는 고유분산을 제거한 축소상관행렬을 분석하게 되고 주성분분석에서는 분산의 공통/고유를 구분하지 않고 모두 분석대상으로 하므로 원상관행렬을 분석하게 된다.

### 부록 3. 문항반응이론(Item response theory)

문항반응이론은 고전검사이론에서와 같이 검사총점에 의하여 분석되는 것이 아니라 각 문항의 속성을 나타내는 문항특성곡선에 의하여 문항을 분석하는 검사이론이다. 문항반응이론은 문항특성불변성 (문항난이도, 문항변별도, 문항추측도가 피험자 집단의 특성에 따라 변화되지 않음)과 피험자능력불변성 (피험자의 능력이 어떤 검사나 문항을 택함으로써 변하는 것이 아니라 고유한 능력수준이 있다는 것, 즉 어려운 검사를 택하든 쉬운 검사를 택하든 능력추정이 같음)의 가정에 근거하고 있다.

문항반응이론을 고전검사이론과 비교하면 몇 가지 점에서 다음과 같은 명백한 차이가 있다.

첫째, 고전검사이론은 검사점수(총점)를 분석하나 문항반응이론은 개별 문항에 대해 분석하므로 고전검사이론에서 가능하지 않은 개별 문항 특성에 대한 평가가 가능하다.

둘째, 고전검사이론은 관찰점수는 진점수와 오차로 구성되어 있다는 가정에서 출발하나 문항반응이론은 문항특성곡선에 근거한다.

셋째, 고전검사이론에서는 진점수를 추정하기 위해 실제로 불가능한 반복측정을 가정한다. 그러나 문항반응이론에서는 그런 가정이 불필요하다.

넷째, 문항난이도, 문항변별도, 문항추측도의 평가에 상이한 수리적 접근이 적용된다. 고전검사이론에서는 문항난이도, 문항변별도, 문항추측도를 각각 정답자 비율, 문항점수와 검사점수와의 상관계수, 정답을 모르고 추측한 피험자의 비율로 정의한다.

이와는 달리 문항반응이론에서는 문항난이도는 문항의 답을 맞힐 확률이 0.5에 대응되는 능력수준의 값, 문항변별도는 문항특성곡선상에 문항난이도를 표시한 점에서의 기울기, 문항추측도는 능력이 전혀 없는 피험자가 문항을 맞힐 확률 등

으로 다른 방법을 적용하여 구한다.

다섯째, 고전검사이론의 경우 문항모수추정치는 실시된 피험자 집단에 종속적인 반면 문항반응이론에서 문항특성은 피험자집단과 무관하여 독립적이다. 고전검사이론에서는 문항 난이도, 변별도 및 특이도 등 문항모수가 각각 정답자 비율, 문항점수와 검사점수와의 상관계수 및 정답을 모르고 추측한 피험자의 비로 정의한다. 따라서 대상 피험자집단의 능력수준과 능력분포에 따라 동일한 문항의 문항특성모수가 달라져서 독립적으로 해석하기가 곤란하나 문항반응이론에서는 문항특성모수의 독립적 해석이 가능하다.

여섯째, 고전검사이론의 경우 피험자와 관련된 능력추정치는 실시된 검사에 종속적이어서 쉬운 검사도구를 이용하면 능력이 과대추정되고 어려운 검사도구를 사용하면 능력이 과소추정된다. 그러나 문항반응이론에서는 검사도구의 특성이 피험자 능력추정치에 영향을 주지 않는다.

일곱째, 고전검사이론의 경우 모든 피험자의 측정오차가 동일하다고 가정하나 실제로는 해당도구가 측정하는 능력수준에 부합하는 집단에 대하여 검사도구의 측정오차가 가장 작으므로 현실과 부합되지 않는 가정에 근거하고 있다. 반면 문항반응이론에서는 측정오차는 해당도구가 측정하는 능력수준에 부합하는 집단에 대하여 검사도구의 측정오차가 가장 작으므로 현실과 부합된다.

여덟째, 고전검사이론의 경우 피험자의 능력과 무관하게 신뢰도가 같다고 간주하나 문항반응이론에서는 피험자에 따라 다르다. 문항반응이론은 해당도구가 측정하는 능력수준에 부합하는 집단에 대하여 검사도구의 신뢰도가 가장 크게 나타나므로 현실과 부합되는 올바른 추정을 수행한다 (성태제, 2009).

결론적으로 앞에서 고전검사이론과 비교한 문항반응이론의 장점을 요약하면 ① 문항모수와 피험자모수가 상호독립적이다. ② 문항모수와 피험자 능력모수가 동일 척도에 놓인다. ③ 피험자 능력 모수가 특정 검사를 위해 표집된 문항들이 아니라 전체 문항 전집과 관련되어 정의된다. ④ 문항차별기능 평가(Differential

item functioning), 컴퓨터 기반 검사(Computerized adaptive testing), 검사 동등화 문제 등에 비교적 손쉬운 적용이 가능하다는 점 등이다(Embreston, 2000).

## 감사의 글

학위논문을 마무리하면서 학위과정에 들어와서 심리학과에서 계량심리 수업을 듣고 과제를 하며 지내던 지난 4년의 기간을 되돌아보게 됩니다. 구강건 강관련 삶의 질 도구를 사용하여 조사를 하고 그 자료를 분석하려고 노력하는 과정에서 어려움을 느낀 저는 보건학 박사 학위를 마친 지 11년 만에 다시 심리학 박사 학위과정을 시작하게 되었습니다. 조금은 의외인 저의 시도를 흔쾌히 허락하여 주신 존경하는 이순목 교수님 덕분에 새로운 발걸음을 땔 수 있습니다. 비록 교수님께서 조기에 퇴임하셔서 전 과정의 지도를 받을 수는 없었지만 학문에 대한 자세와 소중한 가르침은 저의 학문적 삶의 귀감이 되리라 생각합니다.

논문을 쓰는 과정에서 저에게 많은 격려와 도움을 주신 분들께 감사의 인사를 드립니다.

먼저 지도교수이신 조영일 교수님께 감사드립니다. 저의 황당한 질문에 항상 ‘학생이 이해를 못한다면 선생의 탓이다’라고 하시면서 열린 마음으로 받아주시고 미련한 제가 이해할 때까지 여러 각도에서 설명해 주셨습니다. 스승님으로서 귀하신 마음가짐을 항상 깊이 새기려 합니다. 김호 교수님께 감사드립니다. 보건대학원에서의 저의 ‘준 지도교수님’으로서 한 번의 석사와 두 번의 박사학위논문을 기꺼이 지도하여 주셨습니다. 항상 학문적 자극을 주시고 격려하여 주셔서 부족한 제가 나름 맞는 일을 찾아서 지금까지 잘 살아온 것 같습니다. 정선호 교수님께 감사드립니다. PLS-SEM이 무엇인지도 모르면서 대책 없이 시작하였었는데 그나마 여기까지 진행할 수 있었던 것은 전적으로 교수님의 도움 덕분입니다. 서수연 교수님께 감사드립니다. 연구 내용의 전체적인 윤곽을 정하지 못하던 저에게 결정적인 도움을 주셨습니다. 강태훈 교수님께 감사드립니다. 중요한 점들을 깨우쳐 주시고 꼼꼼하게 피드백 해주셔서 논문을

잘 마무리할 수 있었습니다.

소중한 SF-36 데이터를 흔쾌히 제공하여 주신 차보경 교수님께 감사드립니다. 덕분에 제가 실증분석을 진행할 수 있었습니다. 또한 가상 데이터 생성과 처리 과정에 소중한 도움을 주신 최기봉 박사님과 한우리에게 감사합니다.

측정심리 연구실 후배들에게 감사 인사를 전합니다. 저의 학위논문 심사과정에서 누리, 소희와 나라의 세심한 도움을 받았습니다. 또한 따뜻하고 유쾌한 학위과정을 가능하게 해 주었던 지현, 유미, 신용우 선생님과 민형, 유리, 민선, 현서, 송이에게도 감사드립니다. 귀찮은 일도 많았을 터인데 온갖 도움을 아끼지 않았습니다. 고려대 보건통계방 식구들, 은실, 수연, 영선, 다운에게도 감사와 함께 저의 학위과정을 핑계로 지도에 소홀했음에 미안한 마음을 전합니다.

저의 보건대학원 지도교수이신 이승욱 교수님께 감사드립니다. 저의 인생의 방향타를 잡아서 여기까지 이끌어주셨습니다. 고려대 보건정책관리학부 동료 교수님들께 저의 늦은 시도를 이해해주시고 물심양면으로 도와주심에 감사드립니다. 논문이 원활하게 진행되지 않았던 힘든 시기에 쿨하게 통계 연재를 미루어 주신 조병훈 교수님께 감사드리고 두 번째 학위과정 도전의 모범사례를 보여 주신 이신재 교수님께도 감사드립니다. '우리가족같이'팀 식구들, 서울대 보건대학원 통계방 식구들, 대학 동창 친구들, 여고동창 친구들을 비롯하여 허술한 저의 공부를 응원해 주신 많은 분께 감사드립니다.

저의 인생에 가장 든든한 우군인 남편 최진영, 모든 것이 사랑스러운 딸 윤지, 군대에서 힘든 시간을 믿음직하게 잘 보내고 있는 아들 광민. 저의 사랑스러운 가족은 언제나 저의 희망입니다. 연로하셨어도 더 해 줄 것이 없는지 생각하시며 항상 사랑만 주시는 부모님과 지금은 하늘에 계신 시아버님과 가족을 위하여 천상의 기도를 해주시는 천사 같은 시어머님께 감사드립니다. 제가 작으나마 해 낼 수 있는 것이 있다면 그 모든 것은 저를 세상에 나게 하시고 키워주시고 항상 보살펴 주시는 은혜 덕분입니다. 감사합니다.