creative commons

COMMONS DEED

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

dCollection

교육학석사학위논문

IRTree를 활용한 PISA 2018 주관적 안녕감 척도의 극단응답양식 분석

2022년 8월

서울대학교 대학원

교육학과 교육학 전공

오 현 지

IRTree를 활용한 PISA 2018 주관적 안녕감 척도의 극단응답양식 분석

지도교수 박 현 정

이 논문을 교육학 석사 학위논문으로 제출함
2022년 6월

서울대학교 대학원
교육학과 교육학 전공
오 현 지

오현지의 석사 학위논문을 인준함
2022년 8월

위 원 장 _____ 백 순 근 _____ (인)

부위원장 _____ 김 나 영 _____ (인)

위 원 _____ 박 현 정 _____ (인)

국문초록

본 연구의 목적은 PISA(Programme for International Student Assessment) 2018의 주관적 안녕감 척도에 대한 극단응답양식(Extreme Response Style: ERS)과 ERS를 통제한 주관적 안녕감에 대한 잠재변인의 수준을 추정하기에 문항반응나무모형(Item Response Tree model: IRTree model)이 적합한지를 살펴보고, 한국의 ERS 수준이 한국을 제외한 OECD(Organization for Economic Cooperation and Development) 국가들과 어떠한 차이가 있는지 살펴보는 것이다. 또한, ERS 통제 후 한국의 주관적 안녕감에 대한 잠재변인 수준의 변화는 어떠한지 파악하고자 한다. PISA 2018 조사 결과, 한국 학생의 삶의 만족도에 비해 주관적 안녕감이 높은 것으로 나타났다. 선행연구에 따르면, PISA 2018과 같은 자기보고식 검사는 ERS와 같은 응답 편파에 취약하다. 특히 청소년의 안녕감에 대한 선행연구에 따르면, ERS로 인한 응답자 간 응답 차이가 많이 보고 되었다. 또한, PISA 2018의 주관적 안녕감 척도와 같이 4점 척도로 응답범주의 수가 작을 경우, 응답범주의 수가 큰 경우에 비해, ERS의 수준이 증가하고, 이로 인해 측정하고자 하는 구인과 무관하게, 서로 다른 문화적 집단 간 응답 결과의 차이가 발생했다. 따라서, 삶의 만족도에 비해 주관적 안녕감이 높은 이유가 주관적 안녕감 척도 응답 시 ERS가 발생했기 때문인지를 본 연구를 통해 알아보고자 한다. 이를 위해 PISA 2018 주관적 안녕감 척도에 응답한 OECD 32개 국가별로 2,000명의 학생들을 무선표집한 64,000명의 자료를 분석하였다.

연구 문제는 다음과 같다.

연구 문제 1. PISA 2018 주관적 안녕감 데이터에서 ERS와 ERS를 통제

한 주관적 안녕감에 대한 잠재변인의 수준을 추정하기에 IRTree 모형이 적합한가?

연구 문제 2. 한국의 ERS 추정치는 한국을 제외한 OECD 국가들과 어떠한 차이가 있는가?

연구 문제 3. ERS 통제 후 한국의 주관적 안녕감에 대한 잠재변인 수준의 변화는 어떠한가?

본 연구의 주요 결과는 다음과 같다.

첫째, 한국 학생의 PISA 2018 주관적 안녕감 척도 응답에는 ERS가 존재했고, 이를 분석하기에 가장 적합한 모형은 문항반응나무모형(Item Response Tree Model : IRTree model)이었다. 그리고 ERS의 방향성(긍정 또는 부정)이 없다고 가정하는 간명 IRTree 모형은 데이터에 적합하지 않은 것으로 나타났다. 따라서, 주관적 안녕감에 대한 척도 응답에는 ERS의 방향성이 존재한다고 볼 수 있다.

둘째, 등급반응모형(Graded Response Model: GRM)과 IRTree 모형을 통해 추정된 모수 간 상관분석을 통해 IRTree 모형의 신뢰성을 실증하였다. ERS를 통제하기 전과 후의 주관적 안녕감에 대한 잠재변인 추정치 간에는 높은 정적인 상관관계가 존재하였다. 또한, 두 척도의 주관적 안녕감에 대한 잠재변인 추정치 간 상관계수는 ERS 통제 전후에 일정하게 유지되었다. 아울러, 긍정적 정서 척도에 대한 긍정적 ERS 추정치와 부정적 ERS 추정치, 부정적 정서 척도에 대한 긍정적 ERS 추정치의 경우, ERS 통제 후의 주관적 안녕감에 대한 잠재변인 추정치와 높은 정적인 상관을 보였다.

셋째, 한국 학생들이 한국을 제외한 OECD 국가들에 비해 극단응답을 하는 경향이 더 크다는 것을 확인했다. 맨-휘트니 U 검정 및 크루스칼-

왈리스 검정과 그에 따른 사후검정 결과, 한국 학생들은 그 외 국가의 학생들보다 긍정적 정서에 대한 긍정적 ERS, 부정적 정서에 대한 긍정적 및 부정적 ERS 수준이 높았다.

넷째, ERS가 한국 학생들의 주관적 안녕감에 대한 잠재변인의 수준에 유의한 영향을 미친다는 것을 확인했다. 윌콕슨 부호 순위 검정 결과, 긍정적 정서 척도에서는 ERS 통제 후 주관적 안녕감에 대한 잠재변인 추정치가 유의하게 증가했으나, 부정적 정서 척도에서는 ERS 통제 후 주관적 안녕감에 대한 잠재변인 추정치가 유의하게 감소하였다.

다섯째, ERS가 부정적 정서 척도에서 한국 학생들의 주관적 안녕감에 대한 잠재변인 추정치가 한국을 제외한 다른 국가 학생들의 주관적 안녕감에 대한 잠재변인 추정치에 비해 더 큰 것에 기여한다고 추론할 수 있다. 맨-휘트니 U 검정 결과, 부정적 정서 척도에서 ERS을 통제한 후에 한국과 그 외 국가 간 주관적 안녕감에 대한 잠재변인 추정치의 차이가 줄어들었다.

본 연구는 국제 수준의 대규모 자료에서의 ERS의 영향을 경험적으로 확인하였다는 의의를 가진다. 또한, ERS 분석에 대한 IRTree 모형의 효용성을 실증하였다. 추후 연구에서는 응답양식 차이의 원인을 파악하기 위한 실험연구와 질적 면담이 필요할 것이다. 또한, IRTree 모형의 수행력을 파악하기 위한 모의실험 연구를 할 필요가 있다. 그리고 다양한 응답양식을 동시에 추정하는 연구와 가상의 이분문항반응 변환 시의 결측치가 모수 추정에 미치는 영향에 대한 분석이 필요할 것이다.

주요어 : PISA 2018, 주관적 안녕감, 극단응답양식, IRTree 모형

학 번 : 2020-26211

목 차

제 1 장 서론 ····· 1
제 1 절 연구의 필요성 및 목적 ····· 1
제 2 절 연구 문제 ····· 5

제 2 장 이론적 배경 ····· 7
제 1 절 응답양식 ····· 7
1. 응답양식의 종류 ····· 7
2. 응답양식의 원인 ····· 8
3. 응답양식의 영향 ····· 9
제 2 절 기존 문항반응모형의 한계 ····· 11
1. 문항반응이론 ····· 11
2. 등급반응모형 ····· 12
3. 혼합라쉬모형 ····· 14
제 3 절 응답양식 분석을 위한 문항반응모형 ····· 16
1. 다차원명명반응모형 ····· 16
2. 문항반응나무모형 ····· 17
3. 다차원명명반응모형과 문항반응나무모형 간 비교 ····· 25

제 3 장 연구 방법 ····· 27
제 1 절 분석 자료 ····· 27
1. 분석자료 및 분석 대상 ····· 27
2. 분석변수 ····· 28
제 2 절 분석 방법 ····· 30

제 4 장 연구 결과 ····· 36

제 1 절 기술 통계 ····· 36
제 2 절 응답양식 분석 ····· 39
1. 모형적합도 ····· 39
2. 추정된 모수들의 분포 ····· 41
3. 추정된 모수들의 상관관계 ····· 45
제 3 절 국가별 극단응답양식 추정치 비교 ····· 47
1. 한국과 그 외 국가들의 극단응답양식 추정치 비교 ····· 47
2. 국가별 극단응답양식 추정치 비교 ····· 48
제 4 절 극단응답양식 통제 후 주관적 안녕감에 대한 잠재변인 추정치의 변화 ····· 53
1. 한국의 극단응답양식 통제 후 주관적 안녕감에 대한 잠재변인 추정치의 변화 ····· 53
2. 한국과 그 외 국가들의 주관적 안녕감에 대한 잠재변인 추정치 비교 ····· 55

제 5 장 논의 및 결론 ····· 58
제 1 절 요약 및 논의 ····· 58
제 2 절 제한점 및 제언 ····· 62

참고문헌 ····· 64
Abstract ····· 71

표 목 차

<표 2-1> 문항반응이론 로지스틱 모형 공식 ····· 12
<표 3-1> PISA 2018 주관적 안녕감 척도 문항 ····· 28
<표 3-2> IRTree 모형에 따른 범주별 결절 포함 여부 및 가상의 이분문항반응 ····· 32
<표 3-3> IRTree 모형을 적용한 응답범주별 응답 확률 · 33
<표 4-1> 문항별 기술통계 ····· 36
<표 4-2> 국가별 기술통계 ····· 37
<표 4-3> 주관적 안녕감 척도에 대한 모형 적합도 지수 39
<표 4-4> 긍정적 정서 척도에 대한 모형 적합도 지수 ···· 40
<표 4-5> 부정적 정서 척도에 대한 모형 적합도 지수 ···· 40
<표 4-6> 긍정적 정서 척도에 대한 응답 모수 추정치의 기술통계 ····· 42
<표 4-7> 부정적 정서 척도에 대한 응답 모수 추정치의 기술통계 ····· 42
<표 4-8> 전체 학생의 응답 모수 추정치 간 상관관계 ····· 46
<표 4-9> 한국 학생의 응답 모수 추정치 간 상관관계 ····· 46
<표 4-10> 한국과 그 외 국가 간 ERS 추정치 차이에 대한 맨-휘트니 U 검정 결과 ····· 48
<표 4-11> OECD 국가별 ERS 추정치 차이에 대한 크루스칼-왈리스 검정 결과 ····· 48
<표 4-12> OECD 국가별 긍정적 정서 척도의 ERS 추정치 차이에 대한 사후 검정 결과 ····· 51
<표 4-13> OECD 국가별 부정적 정서 척도의 ERS 추정치 차이에 대한 사후 검정 결과 ····· 52

<표 4-14> ERS 통제 전후 한국의 주관적 안녕감에 대한 잠재변인 추정치 차이에 대한 윌콕슨 부호 순위 검정 결과 53
<표 4-15> 한국과 그 외 국가들 간 주관적 안녕감에 대한 잠재변인 추정치 차이에 대한 맨-휘트니 U 검정 결과 ····· 55

그 림 목 차

[그림 2-1] 선형적 나무구조 ····· 18
[그림 2-2] 내재적 나무구조 ····· 18
[그림 2-3] 다분 나무구조 ····· 19
[그림 2-4] 응답의 확실성을 먼저 결정하는 나무구조 ····· 23
[그림 2-5] 응답의 방향을 먼저 결정하는 나무구조 ····· 23
[그림 3-1] 문항반응나무구조 ····· 31
[그림 3-2] 간명 문항반응나무구조 ····· 31
[그림 4-1] 긍정적 정서 척도의 부정적 ERS 추정치에 대한 상자도표 ····· 44
[그림 4-2] 긍정적 정서 척도의 긍정적 ERS 추정치에 대한 상자도표 ····· 44
[그림 4-3] 부정적 정서 척도의 부정적 ERS 추정치에 대한 상자도표 ····· 44
[그림 4-4] 부정적 정서 척도의 긍정적 ERS 추정치에 대한 상자도표 ····· 44
[그림 4-5] 긍정적 정서 척도에서 ERS 통제 전후 한국의 주관적 안녕감에 대한 잠재변인 추정치 상자도표 ····· 54
[그림 4-6] 부정적 정서 척도에서 ERS 통제 전후 한국의 주관적 안녕감에 대한 잠재변인 추정치 상자도표 ····· 54
[그림 4-7] 긍정적 정서 척도에서 ERS 통제 전 주관적 안녕감에 대한 잠재변인 추정치 상자도표 ····· 57
[그림 4-8] 긍정적 정서 척도에서 ERS 통제 후 주관적 안녕감에 대한 잠재변인 추정치 상자도표 ····· 57
[그림 4-9] 부정적 정서 척도에서 ERS 통제 전 주관적 안녕

감에 대한 잠재변인 추정치 상자도표 ····· 57
[그림 4-10] 부정적 정서 척도에서 ERS 통제 후 주관적 안녕감에 대한 잠재변인 추정치 상자도표 ····· 57

제 1 장 서론

제 1 절 연구의 필요성 및 목적

PISA(Programme for International Student Assessment) 2018은 '요즘 자신의 전반적인 삶에 대해 얼마나 만족합니까'라는 질문을 통해 학생들의 삶의 만족도를 조사한다. 이 질문에 대해 0(전혀 만족하지 않는다)부터 10(완벽히 만족한다)까지의 척도로 응답한 값의 평균이 각 국가별 삶의 만족도가 된다. 우리나라 학생들의 삶의 만족도의 평균은 6.52점으로, OECD(Organization for Economic Cooperation and Development) 국가들의 평균인 7.04점에 비해 낮고, 전체 71개국 중에서 65위로 하위권이다(조성민 외, 2019). 이에 따라 우리나라 삶의 만족도가 낮은 이유에 대한 많은 연구들이 수행되어 왔다(김정은 외, 2018; 신승배, 2018; 정혜원 외, 2021).

PISA 2018에서는 학생의 정서와 관련하여 주관적 안녕감(subjective well-being)도 조사하고 있다. 주관적 안녕감은 긍정적 감정과 부정적 감정을 느끼는 빈도에 대해 1(전혀 아니다)부터 4(항상 그렇다)까지의 척도로 응답한 값으로 측정된다. 그리고 주관적 안녕감을 측정하는 척도 중 '행복하다', '즐겁다', '신난다'에 대해 응답한 결과를 결합하여 긍정적 감정 지수를 산출한다. 우리나라 학생들의 긍정적 감정 지수는 0.03으로, OECD 평균인 0.01보다 높다(OECD, 2019). 또한, 주관적 안녕감 척도의 9개 문항 중 8개 문항에서 우리나라 학생들의 극단응답비율(1 또는 4를 선택한 비율)이 OECD 평균보다 높았다. 특히 긍정적 정서에 대한 5개의

문항 중 4개의 문항에서 우리나라 학생들이 4(항상 그렇다)라고 응답한 경우가 OECD 전체 평균보다 높았다(OECD, 2019).

PISA 2018과 같은 자기보고식 검사는 심리적 구인을 측정하는 데 있어 제작하기 간편하고, 실시가 용이하며, 단시간 내 많은 자료를 수집할 수 있기 때문에, 사회과학 분야에서 선호되고 있지만, 심리적 구인을 간접적으로 측정한다는 점에서 측정의 오차를 항상 동반한다. 특히 응답자들은 본인의 의도, 개인적인 성향 및 특성 등에 의하여 응답을 왜곡시키는 경향을 보인다. 이를 응답 편파(response bias)라고 하며, 응답 편파는 타당한 측정을 만족시키려 하는 연구자들에게 매우 큰 위협요인이 된다(손원숙, 2015).

전형적인 응답 편파의 유형 중 하나는 응답양식이다. 응답양식이란, 측정하고자 하는 구인 및 문항의 내용과 관계없이, 문항들에 대한 응답자의 체계적인 응답 경향(Ames & Leventhal, 2021)이다. 대표적인 응답양식으로, 극단응답양식(Extreme Response Style: ERS)이 있다. ERS란, 응답 척도의 가장 높거나 가장 낮은 범주를 지지하는 경향을 의미한다(Ames & Leventhal, 2021). 청소년의 안녕감(well-being)과 관련한 선행연구들에 따르면, 안녕감에 대해 응답할 때 응답자 간 여러 응답양식에서의 차이가 존재했는데, 그중 ERS의 차이가 가장 많이 보고 되었다. 그리고 ERS는 국가, 성별 등의 영향을 받았다(Chen et al., 1995; Clefberg et al., 2015; Gilman et al., 2008; Isabel et al., 2016; Lee et al., 2020).

선행연구들에 따르면, 4점 척도와 같이 응답범주의 수가 작을 경우, 응답범주의 수가 큰 경우에 비해, ERS의 수준이 증가하고(Weijters et al., 2010a), 이로 인해 측정하고자 하는 구인과 무관하게, 서로 다른 문화적 집단 간 응답 결과의 차이가 발생했다(Hui & Triandis, 1989). 따라서, 삶의 만족도에 비해 긍정적 감정 지수가 높은 이유가 주관적 안녕감 척

도 응답 시 ERS가 발생했기 때문인지를 본 연구를 통해 알아보고자 한다.

최근 ERS에 대한 연구에서는 문항반응나무모형 (Item Response Tree model : IRTree model)과 다차원명명반응모형(Multidimensional Nominal Response Model: MNRM)이 가장 활발하게 사용되고 있다. IRTree 모형이란, 응답자의 의사결정 과정을 나무구조로 구체화한 IRT 모형이다(Park & Wu, 2019). 본 연구에서 사용된 PISA 2018 주관적 안녕감 척도에 대한 응답에서 ERS를 분석하는 경우, 1(전혀 아니다)과 2(거의 그렇지 않다) 혹은 3(때때로 그렇다)과 4(항상 그렇다) 중에서 선택하는 단계(문항에 대한 동의 혹은 비동의를 선택하는 단계), 그리고 1과 2 중에서 선택하는 단계(부정적 방향의 극단응답을 선택하는 단계), 3과 4 중에서 선택하는 단계(긍정적 방향의 극단응답을 선택하는 단계)로 구현할 수 있다. 이때, 극단응답을 선택하는 두 단계를 방향에 관계없이 같은 단계로 구현할 수도 있다. 즉, 일반적인 IRTree 모형과 달리, ERS의 방향성(긍정 또는 부정)을 가정하지 않을 수 있으며, 이를 간명 IRTree 모형이라고 한다. 또한, MNRM이란, 명명반응모형에 응답양식을 고려할 수 있는 요소를 추가한 모형이다(Bolt & Johnson, 2009).

IRTree 모형과 MNRM이 ERS 분석에 널리 활용되는 이유는, 다른 모형들이 가지는 한계점을 극복할 수 있기 때문이다. 즉, 혼합라쉬모형을 비롯한 부분점수모형은 기본적으로 응답의 서열성을 가정하기 때문에, 응답양식이 존재할 경우 서열성 가정이 위배될 수 있다는 한계를 가지는 반면(Jeon & De Boeck, 2016) IRTree 모형과 MNRM은 응답의 비서열성을 고려할 수 있다. 특히나 IRTree 모형은 응답자의 의사결정 과정에 대한 가정을 바탕으로, 비교적 자유롭게 모형을 구현할 수 있다는 장점을 가진다(장윤선 외, 2019).

이에 따라, 본 연구에서는 PISA 2018 주관적 안녕감 척도의 ERS와 ERS를 통제한 주관적 안녕감에 대한 잠재변인 수준을 추정하기 위해 IRTree 모형이 적합한지 검증하고자 한다. 이를 위해 자기보고식 척도에 대한 응답을 분석하는 데 널리 활용되어 왔던 GRM, ERS 분석 연구에서 최근 활용되고 있는 MNRM 및 IRTree 모형의 적합도를 비교하고자 한다. 더불어, ERS에서 방향성이 존재하는지를 파악하기 위해, 간명 IRTree 모형의 적합도 또한 함께 비교하였다. 또한, ERS 통제 후 추정된 모수의 신뢰성을 파악하고자, ERS 통제 전후 추정된 모수들의 상관관계를 분석하였다.

그리고 한국 학생들이 실제로 한국을 제외한 OECD 국가들의 학생들보다 극단응답을 하는 경향이 큰지를 살펴보고자 한다. 이를 위해, 맨-휘트니 U 검정(Mann-Whitney U test), 크루스칼-왈리스 검정(Kruskcal-Wallis H test)과 그에 따른 사후비교를 활용하여, 한국과 한국을 제외한 OECD 국가들 간 ERS 수준을 비교하였다.

또한, 한국 학생들의 ERS가 주관적 안녕감에 대한 잠재변인의 수준을 추정하는 데 유의한 영향을 미쳤는지를 파악하기 위해, 윌콕슨 부호 순위 검정(Wilcoxon Singned Rank Test)을 사용하여 ERS 통제 전후 주관적 안녕감 추정치의 크기를 비교하였다. 그리고 맨-휘트니 U 검정을 사용하여 ERS 통제 전후 한국과 한국을 제외한 OECD 국가들 간 주관적 안녕감에 대한 잠재변인 수준의 차이가 유의한지를 검증하였고, 그 차이가 어떻게 변화하였는지를 비교하였다.

제 2 절 연구 문제

본 연구에서는 PISA 2018 주관적 안녕감에 대한 응답을 분석하기 위해 IRTree 모형이 적합한지 검증하고, 국가별 ERS 수준의 차이와 ERS를 통제한 후의 주관적 안녕감에 대한 잠재변인 수준의 변화를 살펴보고자 한다. 본 연구에서 밝히고자 하는 연구 문제는 다음과 같다.

연구 문제 1. PISA 2018 주관적 안녕감 데이터에서 ERS와 ERS를 통제한 주관적 안녕감에 대한 잠재변인의 수준을 추정하기에 IRTree 모형이 적합한가?

1-1. IRTree 모형은 GRM, MNRM, 간명 IRTree 모형에 비해 주관적 안녕감에 대한 잠재변인 수준을 추정하기에 적합한가?

1-2. ERS을 통제하기 전과 후에 추정된 모수들 간 상관은 어떠한가?

연구 문제 2. 한국의 ERS 수준은 한국을 제외한 OECD 국가들과 어떠한 차이가 있는가?

2-1. 한국의 ERS 수준을 한국을 제외한 OECD 국가 전체와 비교할 때, 어떠한 차이가 있는가?

2-2. 한국의 ERS 수준을 한국을 제외한 OECD 개별 국가들과 비교할 때, 어떠한 차이가 있는가?

연구 문제 3. ERS 통제 후 한국의 주관적 안녕감에 대한 잠재변인 수준의 변화는 어떠한가?

3-1. ERS가 한국의 주관적 안녕감에 대한 잠재변인 수준에 유의한 영향을 미치는가?

3-2. ERS 통제 전후 한국과 한국을 제외한 OECD 국가들 간 주관적 안녕감에 대한 잠재변인 수준의 차이는 어떠한가?

제 2 장 이론적 배경

제 1 절 응답양식

1. 응답양식의 종류

응답양식이란, 측정하고자 하는 구인 및 문항의 내용과 관계없이, 문항들에 대한 응답자의 체계적인 응답 경향이다(Ames & Leventhal, 2021). 응답양식에는 극단응답양식(Extreme Response Style: ERS), 중간응답양식(Midpoint Response Style: MRS), 묵인응답양식(Acquiescence Response Style: ARS), 사회적 바람직성 응답양식(Socially Desirable Responding: SDR), 반대응답양식(Disaquiescence Response Style: DRS)이 있다.

ERS란, 응답 척도의 가장 높거나 가장 낮은 범주를 지지하는 경향을 의미하며(Ames & Leventhal, 2021), 많은 선행연구들에서 IRTree 모형을 통해 분석되어 왔다(Ames & Leventhal, 2021; Jeon, & De Boeck, 2016; Plieninge & Heck, 2018; 장윤선 외, 2019; 조영진 & 이현숙, 2022). MRS는 중간 응답 범주를 선택하는 경향을 의미한다(Plieninger & Heck, 2018). 또한, ARS는 긍정적인 응답 범주를 선택하는 경향을 의미한다. 이와 비슷하게, 응답자가 가장 좋게 보이고자, 가장 높은 범주를 선택하는 SDR이 있다. 그리고 DRS는 부정적인 응답 범주를 선택하는 경향을 의미한다(Weijters et al., 2010b).

2. 응답양식의 원인

선행연구를 통해 알려진 응답양식의 원인은 다음과 같다. 첫째, 성격 특성이 응답양식의 원인이 될 수 있다. Austin et al.(2006)에 따르면, ERS를 선호하는 사람들은 외향성과 성실성에서 높은 점수를 얻은 사람들이었다. 이와 비슷하게, 김석호 외(2011)는 응답자의 외향성, 개방성, 친화성, 성실성이 높을수록 ERS의 발생 또한 높아지지만, 정서적 안정성과 ERS는 관계가 없음을 밝혔다. 또한, MRS의 경우, 응답자의 성격특성과는 관계가 없다고 제시하였다. 그러나 성격특성이 응답양식에 미치는 영향을 명확히 분석하는 것은 한계가 있다. 왜냐하면 성격 특성에 대한 자기보고식 검사 점수는 자기표현(self-presentation)욕구에 의해 부풀려질 수 있기 때문이다(Paulhus et al., 1995). 즉, 성격특성과 관련된 검사도구 자체가 응답양식의 영향을 받을 수 있다.

둘째, 문화 특성이 응답양식의 원인이 될 수 있다. Lee & Green(1991)의 연구에 따르면, 한국 학생보다 미국학생이 ERS를 선호했다. 비슷하게, Chen et al.(1995)의 연구에서 일본과 대만 학생들이 북미 학생들보다 MRS를 선호했는데, 개인주의에 대한 지지 정도는 MRS와 부적인 상관을, ERS와 정적인 상관을 보였다. 따라서, ERS와 MRS는 개인주의 및 집단주의 문화와 관련된다는 것을 알 수 있다. 그러나 응답양식과 문화적 차원 간 연관성에 대한 논의는 이론적으로 아직 정립되지 못했다. 왜냐하면, 문화적 특성을 규정짓는 것이 어려울 뿐만 아니라 문항 내용에 따라 응답 성향이 문화별로 차이가 나지 않도록 다양한 주제가 포함되어 있는 설문을 작성하는 것이 불가능하기 때문이다(김석호 외, 2011).

셋째, 응답양식이 나이의 영향을 받는다는 연구 결과들이 있다. Austin et al.(2006)에 따르면, 저연령층과 고연령층은 ERS가 높은 반면에 청장

년층은 상대적으로 낮은 ERS를 보인다고 하였다. 즉, 연령에 따라 ERS의 크기는 U자형으로 달라지는 것으로 나타났다. 그러나 Batchelor & Miao(2016)의 메타연구에서는 20대 초까지 ERS가 증가했다가 이후 감소하는 양상을 보였다.

넷째, 성별이 응답양식에 영향을 준다는 선행연구가 있다. Austin et al.(2006)과 Eid & Rauber(2000)의 연구에서 남성에 비해 여성이 ERS를 더 많이 사용했다. 그러나 Greenleaf(1992)의 연구에서는 성별에 따른 유의한 통계적 차이가 나타나지 않았다.

다섯째, 교육수준이 응답양식의 원인이 될 수 있다. Meisenberg & Williams(2008)에 따르면, 대부분의 국가에서 교육수준이 높을수록, 응답자들은 ARS와 ERS를 덜 사용했다. 그리고 더 젊은 집단이 더 지적 교양이 높다는 점에서 나이와 ERS 간 정적인 관련성 또한 교육수준에 의해 설명될 수도 있다고 주장했다. 이상의 선행연구 결과들을 종합해 볼 때, 응답양식의 원인에 대한 다양한 연구결과들이 존재하지만, 아직까지 응답양식의 원인에 대해 정립된 이론은 없다고 할 수 있다.

3. 응답양식의 영향

리커트척도를 통한 자기보고식 문항반응에서 응답양식을 고려하지 않을 경우, 응답의 비서열성을 고려하지 못하게 된다. 즉, 순차적인 의사결정에 의해 응답범주를 선택하지 않음에도, 응답범주 간 서열성을 가정하고 응답을 분석하게 된다. 이에 따라, 응답자의 잠재적 특성에 대해 편향적 해석을 하게 될 수 있다(Jeon & De Boeck, 2016).

뿐만 아니라, 응답자가 한 검사 내 서로 다른 척도들 사이에서 같은

응답양식을 유지하며, 응답양식이 시간이 지남에도 안정적이라는 선행연구 결과들이 있다(장윤선 외, 2019; Ames & Leventhal, 2021; Weijters 외, 2010). 그리고 성격, 문화, 성별, 나이, 교육수준과 같은 응답자의 개인적인 특성과의 관련성 등에 대한 결과들(김석호 외, 2011; Austin et al., 2006; Batchelor & Miao, 2016; Chen et al., 1995; Eid & Rauber, 2000; Greenleaf, 1992; Lee & Green, 1991; Meisenberg & Williams, 2008)도 응답편파가 응답자 개인의 안정적인 특성이라는 점을 지지한다고 볼 수 있다. 따라서, 응답양식은 후속되는 변인 간의 관계 연구에서도 체계적인 영향을 미칠 수 있다(손원숙, 2015). 비슷한 맥락에서, 응답양식은 성장모형의 모수 추정에도 영향을 미칠 수 있다는 연구 결과도 있다(Soland & Kuhfeld, 2020). 이에 따라 특히 종단연구에서 응답양식으로 인한 개인차를 고려해야 한다는 주장이 있다(Bockenholt, 2017; Ames & Leventhal, 2021).

제 2 절 기존 문항반응모형의 한계

1. 문항반응이론

문항반응이론(Item Response Theory)은 문항의 특성과 피험자의 특성의 함수관계에 의해 피검자가 문항에 정답을 할 확률을 수학적으로 나타내는 측정 모형의 하나이다(백순근 & 채선희, 1998). 문항반응이론은 고전검사이론과 달리, 문항특성의 불변성(the invariance concept of item characteristics)과 피험자 능력 불변성(the invariance concept of ability parameter)을 가지기 때문에 문항 특성 추정과 피험자 능력추정에 널리 사용되고 있다. 문항특성의 불변성이란, 문항 모수들이 피험자 집단의 특성으로부터 비교적 자유로운 것을 의미한다. 또한, 피험자 능력 불변성이란, 어떤 한 순간의 능력을 추정할 때 검사의 특성에 따라 능력이 다르게 추정되지 않는다는 것을 뜻한다(백순근 외, 2007).

문항반응이론모형들은 대개 두 가지 가정을 바탕으로 한다. 첫째, 일차원성 가정(unidimensionality assumption)이란, 검사가 측정하는 내용은 하나의 특성(single trait)이어야 한다는 가정이다. 만약 문항이 여러 특성을 측정하는 경우에는 다차원 문항반응이론 모형을 사용하게 된다. 둘째, 지역독립가정(local independence assumption)이란, 어떤 능력을 가진 피험자의 한 문항에 대한 응답은 다른 문항의 응답에 영향을 주지 않는다는 가정이다. 즉, 어떤 문항의 답을 맞힐 확률과 다른 문항의 답을 맞힐 확률은 상호 독립적이어야 한다.

대표적인 문항반응이론 모형들로는 1모수 모형, 2모수 모형, 3모수 모형이 있다. 첫째, 1모수 모형은 문항모수 중 문항난이도만 고려한 것이

다. 이때 모든 문항의 문항변별도는 1로, 문항추측도는 0으로 가정한다. Rasch 모형은 1모수 로지스틱 모형과 같다. 둘째, 2모수 모형은 문항난이도와 문항변별도를 고려한 모형이다. 셋째, 3모수 모형은 문항난이도, 문항변별도, 문항추측도를 고려한 모형이다. 최근에는 3모수 모형이 가장 널리 사용된다. 각 모형의 로지스틱 모형 공식은 <표 2-1>과 같다(성태제, 2019).

<표 2-1> 문항반응이론 로지스틱 모형 공식

1모수 모형	$P_i(\theta_i) = \frac{1}{1+e^{-1.7(\theta-b)}}$
2모수 모형	$P_i(\theta_i) = \frac{1}{1+e^{-1.7a(\theta-b)}}$
3모수 모형	$P_i(\theta_i) = c+(1-c)\frac{1}{1+e^{-1.7a(\theta-b)}}$

a: 변별도, b: 난이도, c: 추측도

2. 등급반응모형(Graded Response Model)

앞서 서술한 문항반응이론모형들은 이분문항반응을 분석할 수 있으나, 리커트 척도에서 얻어진 응답과 같은 다분문항반응은 분석할 수 없다는 한계를 가진다. 이에 따라 피험자의 반응이 두 종류 이상일 경우에는 다분문항반응이론모형(polytomous item response theory model)을 사용한다. 다분문항반응이론모형은 응답지의 특성에 따라서 크게 두 가지로 구분해 볼 수 있다. 선다형과 같이 반응의 형태가 여러 개이지만 명명적인 특성만을 가진 답지로 구성된 문항에 사용할 수 있는 명명반응모형(Nominal Response Model)과 논술식의 문항과 같이 반응점수의 높고

낮음이 학생의 능력의 정도를 의미하는 경우에 사용할 수 있는 서열반응모형(Ordinal Response Model)으로 구분할 수 있다(박정, 1999).

대표적인 명명반응모형으로는 Bock(1972)의 명명반응모형(Nominal Response Model: NRM)을 들 수 있다. 서열반응모형으로 대표적인 것으로는 Samejima(1969)의 등급반응모형(Graded Response Model: GRM)과 Masters(1982)의 부분점수모형(Partial Credit Model: PCM), Muraki(1992)의 일반화 부분점수모형(Generalized Partial Credit Model: GPCM)을 들 수 있다.

다분문항반응모형이론모형은 단계함수(step function)의 개념에 기초한다. 이는 한 응답범주에서 다음 응답범주로의 전환을 성공할 확률을 수식화한 것이며, <식 2.1>과 같이 2모수 문항반응이론모형을 사용하여 표현할 수 있다.

$$\psi_{ik}(\theta) = \frac{\exp[a_i(\theta - b_{ik})]}{1 + \exp[a_i(\theta - b_{ik})]} \qquad \text{<식 2.1>}$$

이때, $\psi_{ik}(\theta)$는 k번째 단계에서 특정한 θ가 성공할 확률을 의미하며, a_i는 i번째 문항에 대한 변별도를 나타내고, b_{ik}는 i번째 문항의 k번째 단계에서의 난이도를 나타낸다(Penfield, 2014).

GRM의 경우, k번째 단계에서 실패할 경우 k점 미만, 성공할 경우 k점 이상이 된다. 이에 따라 각 전환점의 난이도에 순서가 부여되므로 범주를 묶어서 해석할 수 있다. 또한, 각 문항의 변별도를 다르게 할 수 있기 때문에 서로 다른 문항들에 서로 다른 문항반응함수를 적합시킬 수 있다는 것이 GRM의 장점이다. 그러나 각 단계에서의 변별도는 동일하며, 난이도는 각 단계별로 다른 값을 가진다.

일반적으로 m+1개의 점수 범주를 가진 문항에 대한 GRM은 <식 2.2>와 같이 수식화할 수 있다(Penfield, 2014).

$$Y_i = 0\text{일 때, } P_{i0}(\theta) = 1 - \psi_{i1}(\theta)$$

$$Y_i = \text{j (j>0)일 때, } P_{ij}(\theta) = \psi_{ij}(\theta) - \psi_{ij+1}(\theta) \qquad \text{<식 2.2>}$$

GRM은 서열반응모형으로, 응답자가 '전혀 그렇지 않다'부터 '항상 그렇다'까지의 응답에 대해 일련의 서열성을 바탕으로 의사결정을 한다는 가정을 전제하고 있다. 그러나 응답자가 극단응답양식을 갖고 있을 경우, 이와 같은 서열성 가정은 위배된다. Jeon & De Boeck(2016)에 따르면, 응답척도의 비서열성이 분석에서 무시될 경우, 응답자의 잠재요인 점수는 편향될 수 있다. 따라서, 서열성 가정이 위배되었을 가능성 있는 자기보고식 척도 응답 자료를 분석하기 위해 MNRM과 IRTree와 같은 문항반응이론 모형들이 개발되었다.

3. 혼합라쉬모형 (Mixture Rasch Model)

혼합라쉬모형은 잠재계층분석(latent class analysis)과 라쉬모형을 결합한 것이다. 따라서, 잠재계층분석을 통해 참여자들의 하위집단 간 질적인 차이를 분석할 수 있으며, 동시에 라쉬모형에서 하위집단 내 참여자들 간 양적인 차이를 분석할 수 있다. 즉, 혼합라쉬모형에서는 잠재집단에 따라 문항모수가 변한다(Wetzel et al., 2013).

혼합라쉬모형은 <식 2.3>과 같이 혼합부분점수모형(mixed partial credit model)에 집단 모수를 결합함으로서 정의할 수 있다(Rost, 1991).

$$P(X_{vi}=\chi)=\sum_{g=1}^{G}\pi_g\frac{\exp(\chi\theta_{vg}-\sigma_{ixg})}{\sum_{s=0}^{m}\exp(s\theta_{vg}-\sigma_{isg})} \qquad \text{<식 2.3>}$$

$P(X_{vi}=\chi)$는 응답자 v가 m+1개의 범주를 가진 i번째 문항의 범주 χ를 지지할 확률이며, π_g는 집단의 크기를 의미한다. θ_{vg}는 집단 g에 속한 응답자 v의 특성 수준을 나타내는 응답자 모수이다. 응답 χ가 초과한 모든 임계값들은 문항모수에 축적된다($\sigma_{ixg}=\sum_{s=1}^{\chi}\tau_{isg}$).

잠재계층들이 측정하고자 하는 잠재구인에 대해 동질적이고, 응답양식에서만 차이가 있는지는 앞서 서술한 혼합라쉬모형과 제약된 혼합라쉬모형을 비교함으로써 알아낼 수 있다. 제약된 혼합라쉬모형에서 잠재계층 간 문항모수들은 모두 같다고 제약된다. 즉, $\sigma_{ixg}=\sigma_{ix}$가 성립한다(Wetzel et al., 2013).

Wetzel et al.(2013)이 PISA 2006 데이터를 혼합라쉬모형으로 분석한 결과에 따르면, 잠재구인에 대한 점수가 응답양식에 의해 왜곡되었을 수 있으며, 대다수의 응답자의 응답양식은 서로 다른 잠재구인 간에 일정했다. 그러나 혼합라쉬모형을 비롯한 부분점수모형은 기본적으로 응답의 서열성을 가정하기 때문에, GRM과 마찬가지로, 응답양식이 존재할 경우 서열성 가정이 위배될 수 있다는 한계를 가진다(Jeon & De Boeck, 2016).

제 3 절 응답양식 분석을 위한 문항반응모형

1. 다차원 명명반응모형 (Multimodimensional Nominal Response Model)

<식 2.1>에서의 다분문항반응이론의 단계함수가 NRM에서는 <식 2.4>와 같이 변형된다(Penfield, 2014).

$$\psi_{ik}(\theta) = \frac{\exp(-c_{ik} - a_{ik}\theta)}{1+\exp(-c_{ik} - a_{ik}\theta)} \qquad \text{<식 2.4>}$$

이때 c_{ik}는 <식 2.1>의 b_{ik}와 유사하게 위치 모수로 기능한다. <식 2.1>에서 b_{ik}가 k번째 단계함수의 위치(k번째 단계함수에서 성공확률이 .5일 때의 θ 값)인 것처럼, NRM에서는 $\theta = -c_{ik}/a_{ik}$를 만족할 때, k번째 단계함수에서 성공확률이 .5가 된다. 따라서, NRM은 <식 2.5>와 같이 수식화할 수 있다(Penfield, 2014).

$$Y_i = 0\text{일 때,} \quad P_{i0}(\theta) = \frac{1}{1 + \sum_{k=1}^{m} \exp(c_{ik} + a_{ik}\theta)}$$

$$Y_i = \text{j (j > 0)일 때,} \quad P_{ij}(\theta) = \frac{\exp(c_{ik} + a_{ik}\theta)}{1 + \sum_{k=1}^{m} \exp(c_{ik} + a_{ik}\theta)} \qquad \text{<식 2.5>}$$

그리고 NRM에 응답양식을 고려할 수 있는 요소를 추가하여, MNRM을 <식 2.6>과 같이 수식으로 나타낼 수 있다(Bolt & Johnson, 2009).

$$P(U_j = k|\theta_1, \theta_2, \ldots, \theta_d) = \frac{\exp\left(a_{jk1}\theta_1 + a_{jk2}\theta_2 + \ldots + a_{jkd}\theta_d + c_{jk}\right)}{\sum_{h=1}^{K} \exp\left(a_{jh1}\theta_1 + a_{jh2}\theta_2 + \ldots + a_{jhd}\theta_d + c_{jh}\right)}$$

<식 2.6>

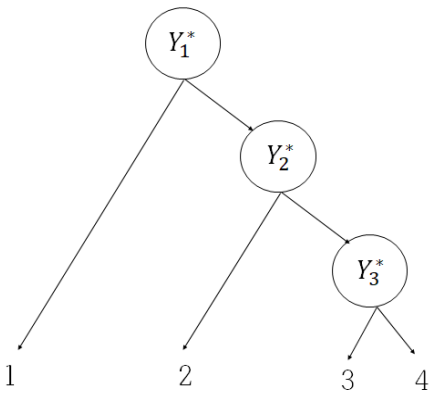
위 식은 응답범주 선택에 영향을 미칠 것이라 가정된 d개의 잠재요인이 주어졌을 때, 응답범주 k를 선택할 확률을 정의하고 있다. d개의 서로 다른 기울기 모수들이 각 문항의 각 응답범주에 대해 정의되며, 각 응답범주에 대한 경향성에 영향을 미친다.

Bolt & Johnson(2009)이 7개의 응답범주를 가진 위스콘신 흡연 의존도 척도(Wisconsin Inventory of Smoking Dependence Motives: WISDM)를 MNRM으로 분석한 결과, 첫 번째 차원은 측정하고자 했던 구인(흡연 의존도)이었고, 두 번째 차원은 ERS라고 볼 수 있는 잠재요인이었다.

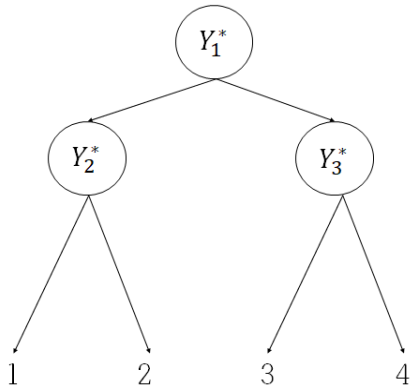
2. 문항반응나무모형 (Item Response Tree model)

IRTree 모형은 문항반응 데이터를 나무 구조로 모형화한 것이다. 나무 구조는 문항의 반응 범주들이나 문항 반응의 기저에 있는 잠재변인들, 혹은 둘 다에 적용될 수 있다(De Boeck & Partchev, 2012). 이 모형은 내재적인 의사결정과정을 하위나무(sub-tree), 내부결절(internal node), 가지를 통해 나무구조로 표현할 수 있다. 나무는 가지를 통해 잎에 도달할 때까지 뻗어나가며, 이때 잎은 단말결절(terminal node)로, 관측된 문

항 반응 범주를 나타낸다. 나무구조에서 원은 결절을, 화살표는 가지를, 잎은 문항 반응 결과를 나타낸다.



[그림 2-1] 선형적 나무구조

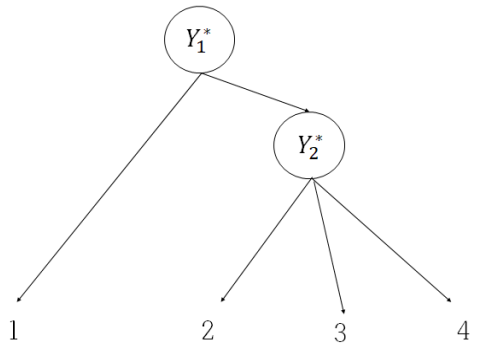


[그림 2-2] 내재적 나무구조

[그림 2-1]은 1부터 4까지의 범주들을 순차적으로 선택하는 경우를 나타낸다. 이와 같은 나무구조를 선형적 나무(linear tree)라고 한다. 선형적 나무구조에서는 각 결절에서 나온 최소한 하나의 가지가 단말결절로 이어져야 한다. 예를 들어, [그림 2-1]은 '전혀 슬프지 않음', '약간 슬픔', '대부분 슬픔', '완전히 슬픔'과 같은 선택지들을 포함하는 단극성 척도(unipolar scale)를 표현할 수 있다.

[그림 2-2]는 두 단계에 걸친 선택 과정이 있는 경우를 나타낸다. 우선 인접한 범주 두 개로 이루어진 집단들 중 하나를 고른 뒤 인접한 범주 둘 중 하나를 최종적으로 선택한다. 이와 같은 나무구조를 내재적 나무(nested tree)라고 한다. 왜냐하면 내부결절로부터 나온 가지가 다른 내부결절로 이어지기 때문이다. [그림 2-2]는 양극성 척도(bipolar scale)를 표현하기 위해 사용될 수 있다. 예를 들어, '완전히 슬픔', '약간 슬픔', '약간 즐거움', '완전히 즐거움'과 같은 척도를 표현할 수 있다. [그림

2-1]과 [그림 2-2]는 이분 나무(binary tree)라고도 할 수 있는데, 두 가지들 사이의 선택만을 포함하고 있기 때문이다.



[그림 2-3] 다분 나무구조

[그림 2-3]은 범주 1을 선택하는 것이 나머지 세 범주를 선택하는 것과 질적으로 다른 경우이다. 예를 들어, 범주 1이 중간 응답인 경우를 표현할 수 있다. 따라서, 범주 1을 선택하지 않을 경우 다음 의사결정이 필요해진다. 이와 같은 경우는 다분 나무(polytomous tree)에 속하는데, 두 개 이상의 가지들 사이의 선택이 포함되기 때문이다(Jeon & De Boeck, 2016)

이와 같이 나무구조가 결정되면, 다분문항반응은 각 의사결정 단계를 의미하는 결절의 가지 방향에 따라 가상의 이분문항반응으로 변환된다. [그림 2-1]~[그림 2-3]에서 Y_k^*는 결절 k에 대한 가상의 이분문항반응을 의미한다. 이 때 이분문항반응의 표현(0 또는 1)은 각 결절의 정의에 따라 달리 사용될 수 있다. 예를 들어, [그림 2-1]의 선형적 나무구조의 범주 1이 '전혀 그렇지 않다', 범주 2가 '그렇지 않다', 범주 3이 '그렇다', 범주 4가 '매우 그렇다'라고 가정해보자. 이때 첫 번째 결절은 부정적 극단반응을 측정하는 결절이 된다. 따라서, '전혀 그렇지 않다'를 선택한 응

답자의 경우 1의 값을, 이외의 경우는 0의 값을 가진다. 즉, 각 결절은 가상의 이분문항반응을 가지는 가상의 문항(pseudo item)이라고 볼 수 있으며, 이와 같은 가상의 이분문항반응은 이분문항반응모형을 통해 각 단계별 의사결정에 영향을 주는 반응양식을 의미하는 잠재변인의 수준을 추정하기 위해 사용된다. [그림 2-1]의 경우, 한 문항에 대한 다분문항반응은 세 개의 가상문항에 대한 가상의 이분문항반응으로 변환될 수 있다.

나무구조에 따라 일부 결절에서는 모든 응답범주가 포함되지 않을 수 있다. [그림 2-1]의 경우, 첫 번째 결절은 4개의 응답범주로 도달하기 위한 경로에 모두 포함되는 반면, 두 번째 결절은 범주 1에 도달하는 경로에는 포함되지 않는다. 이 경우 가상의 이분문항반응을 생성할 때 결측값으로 처리할 수 있다(장윤선 외, 2019).

각 결절에서 생성된 가상의 이분 문항반응의 확률은 2모수 문항반응모형을 통해 <식 2.8>과 같이 정의된다(장윤선 외, 2019).

$$P(Y^*_{ijk}=1)=\frac{\exp\left(a_{ik}(\theta_{jk}-b_{ik})\right)}{1+\exp\left(a_{ik}(\theta_{jk}-b_{ik})\right)}=\Psi\left(a_{ik}(\theta_{jk}-b_{ik})\right) \qquad \text{<식 2.8>}$$

Y^*_{ijk}는 응답자 j의 문항 i에 대한 다분문항반응 Y_{ij}와 나무구조를 바탕으로 새로 생성된 결절 k에 대한 이분문항반응이다. a_{ik}와 b_{ik}는 각각 문항 i의 결절 k에 해당하는 문항변별도와 난이도이다. θ_{jk}는 결절 k에서 측정되는 응답자 j의 잠재변수 추정치를 의미한다. <식 2.8>을 바탕으로 IRTree 모형을 일반화하면 <식 2.9>와 같이 정의된다(Böckenholt & Meiser, 2017).

$$P(Y_{ij}=y)=\prod_{k=1}^{K}P(Y_{ijk}^{*})^{t_{ik}}=\prod_{k=1}^{K}[\frac{\exp(Y_{ijk}^{*}\times a_{ik}(\theta_{jk}-b_{ik}))}{1+\exp(a_{ik}(\theta_{jk}-b_{ik}))}]^{t_{ik}}$$ <식 2.9>

t_{ik}는 결절 k가 응답범주에 도달하기 위한 경로에 포함되는지를 나타내기 위한 이분변수이다. 즉, 결절 k가 포함되지 않는다면, t_{ik}는 0의 값을 가지므로, 응답확률에 영향을 미치지 않는다. IRTree 모형의 결절들은 상호 독립적인 것으로 가정하므로, 응답자의 다분문항반응 확률은 새로 생성된 각 결절의 이분문항반응에 대한 확률의 곱으로 정의된다.

장윤선 외(2019)는 '전혀 그렇지 않다', '그렇지 않다', '그렇다', '매우 그렇다'의 4개 응답범주를 가지는 한국 청소년의 자아존중감, 자아정체감, 우울, 신체증상, 휴대전화 의존도, 다문화 수용도에 대한 문항반응 자료를 [그림 2-2]와 같은 내재적 나무구조로 구현하여, 2모수 문항반응모형을 통해 분석했다. 즉, 첫 번째 결절에서는 긍정적 응답인지, 부정적 응답인지를 선택하고, 이후 두 개의 결절에서는 긍정 혹은 부정하는 정도를 선택한다고 가정하였다. 즉, 첫 번째 결절은 해당 문항이 측정하고자 하는 구인에 대한 개인 특성을 측정하고, 이후의 두 개의 결절은 ERS을 의미하는 잠재변수를 측정하는 것으로 볼 수 있다. 그리고 이후의 두 개의 결절에서 동일한 잠재변인을 측정한다는 제약을 추가한 간명 IRTree 모형과 GRM을 통해 분석한 결과와 비교하였다. 세 모형의 적합도를 비교한 결과, IRTree 모형의 적합도가 가장 높았다. 따라서, ERS를 구인과 다른 별개의 잠재변인으로 분리하는 것이 서열척도 자료 분석에 더 적합하며, ERS에는 긍정과 부정이라는 방향성이 존재한다는 결론을 도출하였다. 뿐만 아니라, 부정적 및 긍정적 ERS 추정치 간 상관은 구인에 관계없이 모두 정적 상관이 있는 것으로 나타났다. 따라서, ERS는 심리적 구인의 특성과 무관한 개인의 일관적 특성임이 확인하였다.

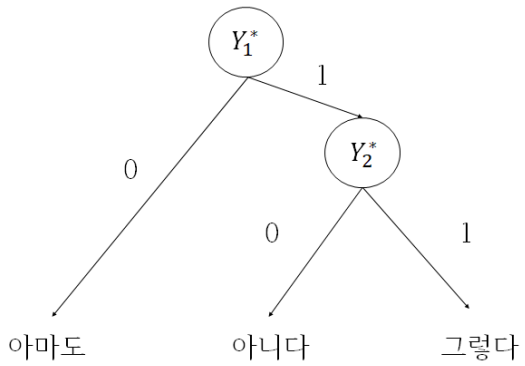
조영진, 이현숙(2022)는 유럽, 아시아, 북미 청소년의 과학에 대한 흥미와 자신감에 대한 문항반응 자료를 [그림 2-2]와 같은 내재적 나무구조로 분석했다. 이 경우에도 응답범주는 '전혀 그렇지 않다', '그렇지 않다', '그렇다', '매우 그렇다'의 4개 응답범주였다. 나무구조 또한 장윤선 외(2019)의 간명 IRTree 모형과 동일했다. 즉, 첫 번째 결절은 응답양식이 통제되었을 때 응답자의 척도에 대한 동의 정도를 측정했으며, 두 번째 결절은 ERS 모수를 측정했다. 분석 결과, GRM보다 IRTree모형의 적합도가 더 높아 자료 내 ERS가 존재함을 알 수 있었다. 또한, 국가, 성별, 나이, 성취수준에 따른 ERS의 차이가 나타났으며, ERS가 클수록 ERS 통제 전후에 응답자 모수의 변화가 커졌다고 보고했다.

일반적인 IRTree 모형은 각 결절이 개별적인 잠재요인을 가지며, 응답척도 자체가 다차원적임을 가정한다. 이는 측정하고자 하는 잠재요인의 점수가 어떻게 결정될 것인가 하는 문제를 발생시킨다. 그래서 Jeon & De Boeck(2016)은 쌍요인(bifactor) IRTree 모형를 제안했다. 이는 결절들이 부분적으로만 다르고, 공통의 잠재요인을 공유한다고 가정하는 것이다. 예를 들어, 우울 척도에서 공통의 잠재요인은 우울이 된다. 공통의 잠재요인이 주어졌을 때, 각 결절의 개별 잠재요인이 서로 독립적이라고 가정하면, IRTree 모형은 <식 2.10>과 같이 쌍요인 구조로 변형될 수 있다.

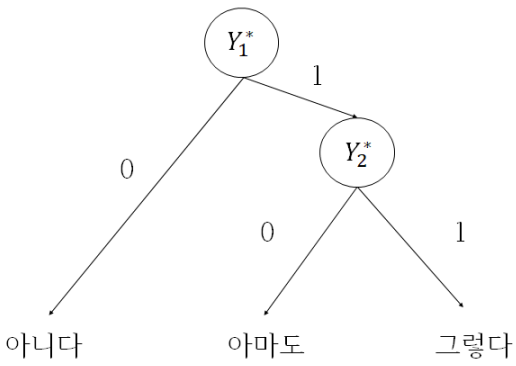
$$P(Y^{*}_{ijk}=1|\theta_{jk})=g^{-1}(\alpha^{g}_{ik}\theta^{g}_{j}+\alpha_{ik}\theta_{jk}+\beta_{ik}) \qquad \text{<식 2.10>}$$

α^{g}_{ik}는 문항 i의 결절 k에서의 기울기 모수이고, 이는 응답자 j에 대한 공통의 잠재요인인 θ^{g}_{j}와 관련되어 있다. 또한, α_{ik}는 문항 i의 결절 k에

대한 잠재요인 θ_{jk}에 대한 개별적인 기울기 모수이다. 이때, K+1개의 잠재요인들이 특정되고, 서로 독립적이라고 가정된다.



[그림 2-4] 응답의 확실성을 먼저 결정하는 나무구조



[그림 2-5] 응답의 방향성을 먼저 결정하는 나무구조

Jeon & De Boeck(2016)은 '아니다', '아마도', '그렇다'의 세 응답범주로 구성된 언어 공격성 척도 응답 자료를 IRTree 모형와 쌍요인나무모형으로 분석했다. 이때, '아마도'라는 응답이 언어 공격성을 인정하는 응답으로 볼 것인지에 따라 다른 나무구조를 구성할 수 있다. 예를 들어, 첫 번째 결절에서 응답자가 확실한 정도를 선택하고, 두 번째 결절에서 응답의 방향성(긍정 혹은 부정)을 선택한다고 가정할 경우, [그림 2-4]와 같은 나무구조로 표현될 수 있다. 반면에, 첫 번째 결절에서 응답의 방향성(긍정 혹은 부정)을 선택하고, 두 번째 결절에서 응답의 확실한 정도를 선택한다고 가정할 경우, [그림 2-5]와 같은 나무구조로 표현될 수 있다.

모형 적합도 분석 결과, [그림 2-5]와 같은 구조가 더 적합했다. 또한, 두 가지 의사결정 과정이 매우 약한 상관관계를 가지고 있다는 점에서, 세 응답범주가 서열적이지 않다고 판단하였다. 그리고 기존의 IRTree 모

형과 더불어, IRTree 모형에서 각 결절의 기울기 및 절편이 동일하다고 가정한 간명 모형, 단일한 잠재변인을 가정하는 일차원 나무모형, 공통의 잠재요인과 각 결절별 잠재요인을 가정하는 쌍요인나무모형 간 적합도를 비교한 결과, 쌍요인 나무모형이 가장 적합도가 높았다고 보고했다. 따라서, 이 척도에서 결절별 측정모수는 변하며, 척도는 다차원적이지만, 공통의 잠재요인이 측정될 수 있다고 결론 지었다. 뿐만 아니라, GRM 분석 결과와의 비교를 통해, 응답척도의 비서열성이 분석에서 무시될 경우, 응답자의 잠재요인 점수는 편향될 수 있음을 주장했다.

Jeon & De Boeck(2019)은 ERS를 측정하기 위한 표준 IRTree 모형(standard IRTree model)의 두 가지 가정을 없앤 일반 IRTree 모형(general IRTree model)을 제안했다. 두 가지 가정이란, 방향적 불변성(directional invariance)과 내용적 불변성(content invariance)이다. 방향적 불변성이란, 높은 ERS를 보이는 응답자는 그들이 동의하는지와 관계없이, 더 극단적인 응답을 선택할 것이라는 가정이다. 내용적 불변성이란, 서로 다른 내용을 질문하는 문항들 간 극단적으로 응답하려는 행동이 변하지 않는다는 가정이다. 즉, 표준 IRTree 모형은 앞서 서술한 간명 IRTree 모형에 내용적 불변성 가정을 추가한 모형이라고 볼 수 있다. Jeon & De Boeck(2019)이 Open Source Psychometrics Project에 있는 성격척도를 활용한 데이터를 분석한 결과, 일반화 IRTree 모형이 표준 IRTree 모형보다 적합도가 더 높았다. 또한, 응답자들은 응답의 방향과 관계없이 극단응답범주를 선택하려는 경향(방향적 불변성)을 보였으나, 문항의 내용과 상관없이 극단응답범주를 선택하려는 경향(내용적 불변성)은 다소 적었다.

이와 같은 선행연구들의 다양한 IRTree 모형 활용 사례에서 알 수 있듯이, IRTree 모형은 의사결정 단계에 대한 가정을 바탕으로 비교적 자

유롭게 모형을 구현할 수 있다. 따라서, IRTree 모형은 응답양식에 대한 연구자의 가정을 검증하는 확인적 접근의 측면이 강하다. 또한, IRTree 모형은 응답자와 문항 특성을 동시에 고려할 수 있다는 점에서 기존에 제안된 확장된 PCM과 다차원 IRT 모형과 차이가 있다(장윤선 외, 2019).

3. 다차원명명반응모형과 문항반응나무모형 간 비교

앞서 서술했듯이, IRTree 모형과 MNRM은 응답의 비서열성을 고려할 수 있다. 하지만 두 모형은 응답양식과 측정하는 구인 간 관계에 대한 가정을 다르게 하고 있다. MNRM은 보상적 다차원 모형으로서 응답양식이 측정하고자 하는 구인을 보상할 수 있다고 가정한다. 보상적 다차원 모형이란, 한 개 이상의 차원(들)에 대한 응답자의 능력이 다른 잠재변인(들)에 대한 응답자의 능력을 보상하는 경우를 모형화한 것이다(de Ayala, 2009). 1점(매우 아니다)부터 4점(매우 그렇다)까지의 리커트 척도에 대한 응답을 MNRM으로 분석한다고 가정해보자. 응답자가 응답양식과 측정하고자 하는 구인 둘 중 하나만 높은 수준으로 가지고 있어도, 4점(매우 그렇다)을 선택하게 된다.

반면에, IRTree 모형은 응답양식과 측정하고자 하는 구인이 서로 다른 응답 과정에서 독립적으로 기능한다고 가정한다는 차이점이 있다(Zhang & Wang, 2020). 이와 같은 비보상적 다차원 모형의 경우, 응답자의 잠재구인(들)에 대한 능력(들)이 다른 잠재구인에 대한 능력(들)을 보상하지 못한다(de Ayala, 2009). 1점(매우 아니다)부터 4점(매우 그렇다)까지의 리커트 척도에 대한 응답을 IRTree 모형으로 분석한다고 가정해보자.

IRTree 모형의 경우, 오로지 ERS에 의해서만 극단응답(1점과 4점)을 선택하는 결정이 이루어지기 때문에, 아무리 측정하고자 하는 구인이 높더라도, ERS가 낮은 응답자의 경우, 4점을 선택하지 않게 된다.

뿐만 아니라, IRTree 모형은 의사결정 단계에 대한 가정을 바탕으로 비교적 자유롭게 모형을 구현해낼 수 있다. 따라서, 응답자와 문항의 특성을 동시에 고려할 수 있고(장윤선 외, 2019), 다양한 응답양식을 동시에 고려하여 측정하고자 하는 잠재구인으로부터 분리해낼 수 있는 유연성을 가지고 있다는 점(Ames & Leventhal, 2021)에서 MNRM과 차이가 있다.

제 3 장 연구방법

제 1 절 분석 자료

1. 분석자료 및 분석 대상

본 연구에서는 OECD에서 제공하는 PISA 2018 자료를 활용하였다. PISA는 3년마다 참여국의 만 15세 학생들의 사회경제적 삶을 위한 주요 지식 및 능력의 습득 정도를 측정하는 평가이다(OECD, 2019). 본 연구에서 활용한 응답 데이터는 학생의 안녕감(students' well-being) 문항에 대한 응답 데이터이다. PISA 2018에서는 학생의 안녕감을 삶의 만족도(eudemonia)와 주관적 안녕감으로 나누어 측정하고 있다. 이때, 삶의 만족도의 경우, 측정을 위해 사용된 문항이 총 3개로 문항반응모형으로 분석하기에 적합하지 않아 분석에서 제외하였다.

연구 대상으로는 주관적 안녕감 척도에 응답한 OECD 32개 국가들을 선정하였다. 선정된 국가들의 응답자 수는 호주 6,206명, 캐나다 1,9304명, 스위스 4,961명, 칠레 6,452명, 콜롬비아 6,524명, 체코 6,392명, 독일 3,963명, 덴마크 6,396명, 스페인 3,1841명, 에스토니아 4,998명, 핀란드 5,206명, 프랑스 528명, 영국 12,400명, 그리스 5,854명, 헝가리 4,805명, 아일랜드 5,220명, 아이슬란드 2,829명, 이탈리아 10,252명, 일본 5,866명, 한국 6,489명, 리투아니아 6,250명, 룩셈부르크 4,657명, 라트비아 4,880명, 멕시코 5,508명, 네덜란드 3,687명, 폴란드 5,347명, 포르투갈 5,388명, 슬

로바키아 5,280명, 슬로베니아 5,830명, 스웨덴 4,950명, 터키 6,460명, 미국 4,442명으로, 표본 크기가 국가별로 달랐다. 따라서, 표본 크기의 영향을 통제하고자 각 국가별로 2,000명의 학생들을 무선표집하여 64,000명의 자료를 분석하였다.

2. 분석 변수

주관적 안녕감은 긍정적 정서와 부정적 정서를 느끼는 빈도에 대해 'Never(전혀 그렇지 않다)', 'Rarely(거의 그렇지 않다)', 'Sometimes(때때로 그렇다)', Always(항상 그렇다)'의 4점 척도로 측정되었다. 이를 측정하는 문항은 총 9문항으로 아래 <표 3-1>과 같다.

<표 3-1> PISA 2018 주관적 안녕감 척도 문항

문항 번호	문항 내용
공통	Thinking about yourself and how you normally feel: how often do you feel as described below?
ST186Q05HA	Happy
ST186Q06HA	Scared*
ST186Q07HA	Lively
ST186Q10HA	Miserable*
ST186Q09HA	Proud
ST186Q02HA	Afraid*
ST186Q01HA	Joyful
ST186Q08HA	Sad*
ST186Q03HA	Cheerful

*: 부정적 정서를 묻는 문항

이 때, 긍정적 정서를 묻는 5문항과 부정적 정서를 묻는 4문항을 별도로 분석하였다. 왜냐하면, 9문항이 모두 주관적 안녕감을 측정하는 문항들이기는 하지만, PISA 2018에서 긍정적 정서와 부정적 정서를 나누어 해석하고 있으며(OECD, 2019), 부정적 정서와 긍정적 정서는 독립적인 차원이라는 점이 여러 연구(Diener & Emmons, 1985; Warr et al., 1983;. Watson & Clark, 1997)에 의해 밝혀졌기 때문이다.

제 2 절 분석 방법

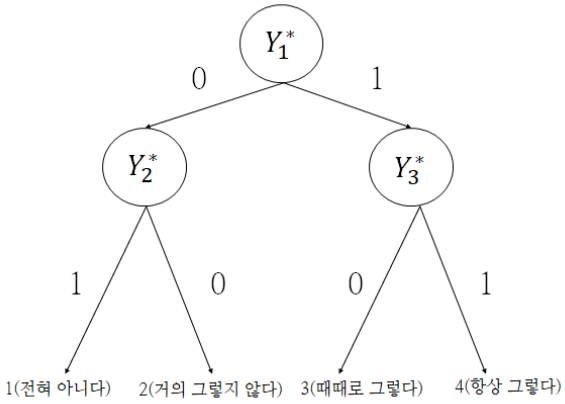
본 연구는 ERS를 통제한 주관적 안녕감에 대한 잠재변인의 수준을 추정하기에 IRTree 모형이 적합한지를 살펴보고, 한국의 ERS 수준이 한국을 제외한 OECD 국가들과 어떠한 차이가 있는지 살펴보는 것이다. 또한, ERS 통제 후 한국의 주관적 안녕감에 대한 잠재변인 수준의 변화는 어떠한지 파악하고자 한다. 이를 위한 분석 방법은 아래와 같다.

첫째, PISA 2018 주관적 안녕감 척도에 대한 응답을 분석하기 위해 IRTree 모형이 가장 적합한지를 알아보기 위해, GRM, MNRM, IRTree 모형, 간명 IRTree 모형의 적합도를 비교하고자 한다. 자기보고식 척도를 분석하기 위해 널리 활용되어 왔던 GRM은 응답의 서열성을 가정하여 응답양식에 대한 고려 없이, 다분문항반응자료를 분석한다. 반면, IRTree 모형과 MNRM은 응답의 서열성을 가정하지 않고, ERS과 측정하고자 하는 구인을 동시에 추정한다.

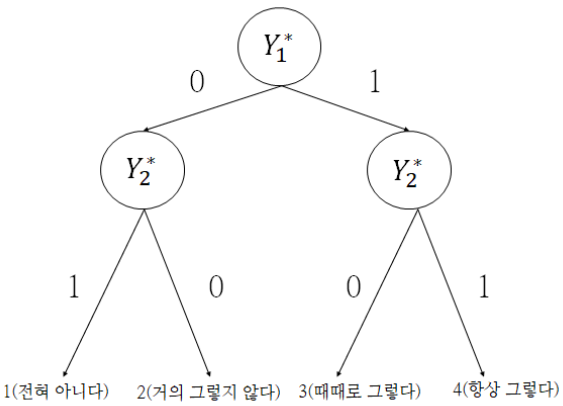
IRTree 모형으로 자료를 분석하기 위해 다분문항반응을 나무구조의 결절에 따른 가상의 이분문항반응자료로 변환해야 한다. 이 연구에서 분석하는 자료는 4점 척도로, [그림 3-1]과 같은 내재적 문항반응나무구조를 가정하여 IRTree 모형을 적용하였다. 이때, Y_k^*는 결절 k에 대한 가상의 이분문항반응이다. 이 모형을 통해 주관적 안녕감을 나타내는 잠재변수 추정치(θ_1), 부정적 ERS를 나타내는 잠재변수 추정치(θ_2), 긍정적 ERS를 나타내는 잠재변수 추정치(θ_3)의 총 3개의 모수를 추정했다.

간명 IRTree 모형이란, ERS가 문항반응의 방향성(긍정 또는 부정)과 무관하다는 가정을 바탕으로, [그림 3-2]와 같이, [그림 3-1]에서의 θ_2와 θ_3가 동일한 잠재변인이라는 제약을 추가한 모형이다. 네 모형의 적합도

를 비교함으로써 자료에 ERS가 존재하고, ERS로 인해 자료의 서열성이 위배되며, 이 자료를 분석하기 위해 가장 적절한 모형이 IRTree 모형임을 보이고자 한다. 또한, ERS 통제 후 추정된 모수의 신뢰성을 파악하고자, ERS 통제 전후 추정된 모수들의 상관관계를 분석하였다.



[그림 3-1] 문항반응나무구조



[그림 3-2] 간명 문항반응나무구조

기본 IRTree 모형과 간명 IRTree 모형에서, 응답자 j의 문항 i에 대한

다분문항반응(Y_{ij})에 따른 각 결절 포함 여부에 대한 이분 변인(t_{ik}), 가상의 이분문항반응(Y^*_{ijk})은 <표 3-2>와 같다. 주관적 안녕감을 측정하는 첫 번째 결절의 경우, 모든 응답범주로 가는 경로에 포함되기 때문에, t_{i1}은 (1, 1, 1, 1)으로 정의된다. 부정적 ERS를 측정하는 두 번째 결절의 경우, '전혀 아니다'와 '거의 그렇지 않다'로 가는 경로에는 포함되지만, '때때로 그렇다'와 '항상 그렇다'로 가는 경로에는 포함이 되지 않기 때문에, t_{i2}은 (1, 1, 0, 0)이 된다. 반면에, 긍정적 ERS를 측정하는 세 번째 결절의 경우, '전혀 아니다'와 '거의 그렇지 않다'로 가는 경로에는 포함되지 않지만, '때때로 그렇다'와 '항상 그렇다'로 가는 경로에는 포함되기 때문에, t_{i3}은 (0, 0, 1, 1)이 된다.

<표 3-2> IRTree 모형에 따른 범주별 결절 포함 여부 및 가상의 이분문항반응

	Y_{ij}	t_{i1}	t_{i2}	t_{i3}	Y^*_{ij1}	Y^*_{ij2}	Y^*_{ij3}
IRTree 모형	1	1	1	0	0	1	NA
	2	1	1	0	0	0	NA
	3	1	0	1	1	NA	0
	4	1	0	1	1	NA	1
간명 IRTree 모형	1	1	1		0	1	
	2	1	1		0	0	
	3	1	1		1	0	
	4	1	1		1	1	

Y_{ij}: 응답자 j의 문항 i에 대한 다분 문항반응

t_{ik}: 응답자 j의 의사결정 경로에서 결절 k 포함 여부

Y^*_{ijk}: 응답자 j의 문항 i에서 결절 k에 대한 가상의 이분문항반응

예를 들어, ‘항상 그렇다’를 선택한 경우, 첫 번째 결절에서의 문항반응(Y^*_{ij1})은 1의 값을 가지며, 두 번째 결절은 경로에 포함되지 않기 때문에, 이에 대한 문항반응(Y^*_{ij2})은 결측값으로 처리된다. 세 번째 결절의 경우, ‘항상 그렇다’가 극단 응답에 해당하기 때문에 해당하는 문항반응(Y^*_{ij3})은 1의 값을 가지게 된다. 2PL IRT모형을 적용하여 구한 각 응답범주별 응답 확률은 <표 3-3>과 같다.

<표 3-3> IRTree 모형을 적용한 응답범주별 응답 확률

	Y_{ij}	응답 확률
IRTree 모형	1	$P(Y_{ij}=1)=[1-\Psi\{a_{i1}(\theta_{j1}-b_{i1})\}]\times\Psi\{a_{i2}(\theta_{j2}-b_{i2})\}$
	2	$P(Y_{ij}=2)=[1-\Psi\{a_{i1}(\theta_{j1}-b_{i1})\}]\times[1-\Psi\{a_{i2}(\theta_{j2}-b_{i2})\}]$
	3	$P(Y_{ij}=3)=\Psi\{a_{i1}(\theta_{j1}-b_{i1})\}\times[1-\Psi\{a_{i3}(\theta_{j3}-b_{i3})\}]$
	4	$P(Y_{ij}=4)=\Psi\{a_{i1}(\theta_{j1}-b_{i1})\}\times\Psi\{a_{i3}(\theta_{j3}-b_{i3})\}$
간명 IRTree 모형	1	$P(Y_{ij}=1)=[1-\Psi\{a_{i1}(\theta_{j1}-b_{i1})\}]\times\Psi\{a_{i2}(\theta_{j2}-b_{i2})\}$
	2	$P(Y_{ij}=2)=[1-\Psi\{a_{i1}(\theta_{j1}-b_{i1})\}]\times[1-\Psi\{a_{i2}(\theta_{j2}-b_{i2})\}]$
	3	$P(Y_{ij}=3)=\Psi\{a_{i1}(\theta_{j1}-b_{i1})\}\times[1-\Psi\{a_{i2}(\theta_{j2}-b_{i2})\}]$
	4	$P(Y_{ij}=4)=\Psi\{a_{i1}(\theta_{j1}-b_{i1})\}\times\Psi\{a_{i2}(\theta_{j2}-b_{i2})\}$

자료의 다분 문항반응자료를 가상의 이분 문항반응자료로 변환하기 위해서는 기존 자료와 나무구조를 반영한 사상행렬(mapping matrix)을 정의해야 한다. 사상행렬의 행과 열은 각각 응답범주와 IRTree 모형에 의해 추정되는 잠재변인(나무구조의 결절)을 의미한다. 즉, 각 결절의 가지

방향에 의해 정의되는 가상의 이분문항반응을 정의하는 행렬이다(장윤선 외, 2019). 이 연구에서 정의한 IRTree 모형의 사상행렬은 <식 3.1>과 같으며, 간명 IRTree 모형의 사상행렬은 <식 3.2>와 같다.

$$\text{IRTree 모형의 사상행렬} = \begin{bmatrix} 0 & 1 & NA \\ 0 & 0 & NA \\ 1 & NA & 0 \\ 1 & NA & 1 \end{bmatrix} \qquad \text{<식 3.1>}$$

$$\text{간명 IRTree 모형의 사상행렬} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \qquad \text{<식 3.2>}$$

가상의 이분문항반응 자료는 R 패키지 irtrees(De Boeck & Partchev, 2012)의 dendrify 함수를 사용하여 생성되었다. 또한 각 문항반응모형의 모수는 R 패키지 mirt(Chalmers, 2012)를 사용하여 분석되었다. 그리고 각 모형의 적합도 비교를 위해 대표적인 정보지수인 AIC(Akaike Information Criterion), BIC(Bayesian Inforamtion Criterion), SABIC(Sample-size Adjusted BIC)를 사용하였다. 문항 수준의 적합도와 피험자 수준의 적합도를 비교하지 않은 이유는, IRTree 모형에서 가상의 이분문항을 만듦으로써, 문항 간 직접적 비교가 불가능해졌고, IRTree 모형의 경우, 각 피험자의 결측치가 필수적으로 생성되기 때문에 피험자 수준의 적합도를 구할 수 없기 때문이다.

둘째, 모형을 통해 추정된 ERS의 수준에 대해 국가 간 비교를 하고자 한다. 이를 위해 우선 추정된 모수들의 정규성을 샤피로-윌크 검정(Shapiro-Wilk test)과 앤더슨-달링 검정(Anderson-Darling test)을 통해 확인하였다. 그 결과, 모든 분포들이 정규성을 만족하지 않는 것으로 나타났다. 이에 따라, 한국과 그 외 국가들 간 추정된 ERS의 수준에 대한

비교를 위해서 맨-휘트니 U 검정(Mann-Whitney U test)을 사용하였다. 또한, 각 국가별로 추정된 ERS 수준의 차이가 있는지 알아보기 위하여 크루스칼-왈리스 검정(Kruskcal-Wallis H test)을 실시하였다. 그리고 이에 대한 사후검정으로서, Hochberg 방법과 Benjamini-Hochberg 방법을 둘 다 사용하여 그 결과를 비교하였다. 두 방법을 모두 사용한 이유는 FWER(Family-Wise Error Rate)을 통제하는 Hochberg 방법이 검사의 수가 증가할 수록, 제 2종의 오류가 증가한다는 단점이 있기 때문이다(Verhoeven et al., 2005). 이에 따라, FDR(False Discovery Rate)을 통제함으로써, 전체 1종의 오류 발생을 고려하는 Benjamini-Hochberg 방법(Benjamini & Hochberg, 1995)도 동시에 사용하여 그 결과를 비교함으로써, 보다 엄밀한 검정을 수행하였다.

셋째, 한국 학생들에게서 ERS 통제 후 주관적 안녕감에 대한 잠재변인 수준의 변화가 유의한지 알아보기 위해, 윌콕슨 부호 순위 검정(Wilcoxon Singned Rank Test)을 사용하였다. 또, 한국과 그 외 국가들 간 추정된 주관적 안녕감에 대한 잠재변인의 수준에 대한 비교를 위해서 맨-휘트니 U 검정(Mann-Whitney U test)을 사용하였다.

제 4 장 연구결과

제 1 절 기술통계

OECD 국가별 무선으로 추출된 2,000명의 표본을 대상으로 각 문항별로 구한 기술통계치는 <표 4-1>과 같다. Kline(2005)이 제시한 기준에 따라, 모든 문항들의 왜도의 절댓값이 3보다 작고, 첨도의 절댓값이 10보다 작으므로, 자료는 정규분포를 따르고 있다고 볼 수 있다.

<표 4-1> 문항별 기술통계

문항 번호	문항 내용	평균	표준편차	왜도	첨도
ST186Q05HA	Happy	3.310	.669	-.747	.634
ST186Q06HA	Scared*	2.225	.756	.145	-.368
ST186Q07HA	Lively	3.118	.773	-.647	.105
ST186Q10HA	Miserable*	2.267	.806	.089	-.569
ST186Q09HA	Proud	2.849	.779	-.393	-.120
ST186Q02HA	Afraid*	2.480	.834	-.026	-.572
ST186Q01HA	Joyful	3.287	.698	-.776	.542
ST186Q08HA	Sad*	2.488	.734	-.093	-.304
ST186Q03HA	Cheerful	3.268	.722	-.807	.522

*: 부정적 정서를 묻는 문항

<표 4-2> 국가별 기술통계

	긍정적 정서				부정적 정서			
국가	평균	표준편차	왜도	첨도	평균	표준편차	왜도	첨도
한국	3.193	.641	-.636	.067	2.465	.686	-.064	-.358
일본	3.074	.561	-.780	1.396	2.695	.598	-.369	.368
호주	3.227	.569	-.997	1.396	2.334	.575	-.002	-.072
캐나다	3.124	.581	-.696	.847	2.362	.656	.070	-.185
멕시코	3.386	.494	-.890	1.067	2.239	.609	.103	-.032
미국	3.133	.604	-.546	.269	2.400	.634	-.052	-.211
칠레	3.246	.544	-.637	.197	2.235	.597	.103	-.098
콜롬비아	3.294	.513	-.766	.756	2.383	.514	.175	.222
스위스	3.259	.513	-.876	1.467	2.354	.551	-.006	.098
체코	3.103	.580	-.605	.570	2.511	.597	-.188	.017
독일	3.200	.529	-.719	.699	2.384	.551	.008	-.065
덴마크	3.305	.476	-.777	1.575	2.136	.598	.302	-.248
스페인	3.316	.527	-.844	1.225	2.450	.520	-.148	.095
에스토니아	3.076	.534	-.442	.139	2.424	.588	.160	.025
핀란드	3.085	.515	-.695	1.571	2.322	.559	.048	-.064
프랑스	3.199	.513	-.870	1.404	2.366	.582	.036	.032
영국	3.033	.554	-.521	.725	2.534	.604	-.182	-.175
그리스	3.201	.532	-.775	1.532	2.417	.572	.044	.076
헝가리	3.229	.590	-.892	.863	2.302	.615	.131	-.147
아일랜드	3.176	.498	-.442	.398	2.331	.548	-.069	-.153
아이슬란드	3.178	.583	-.751	1.351	2.210	.598	.060	-.197
이탈리아	2.979	.507	-.490	.571	2.357	.547	.118	.224
리투아니아	3.212	.542	-.625	.568	2.175	.627	.194	-.162
룩셈부르크	3.216	.577	-.814	.909	2.426	.581	-.021	.034
네덜란드	3.235	.460	-.565	.738	2.300	.557	-.066	-.202
폴란드	3.120	.576	-.621	.469	2.305	.573	.360	.409
포르투갈	3.258	.480	-.687	1.005	2.453	.507	-.114	.392
슬로바키아	3.059	.575	-.644	.624	2.455	.586	.059	.175
슬로베니아	2.906	.550	-.394	.667	2.314	.552	.462	.459
스웨덴	3.132	.567	-.763	1.227	2.376	.609	-.038	-.083
터키	3.076	.644	-.705	.459	2.263	.594	.328	.222
전체	3.166	.557	-.697	.816	2.365	.594	.045	-.041

주: 한국을 제외한 국가는 아시아, 오세아니아, 아메리카, 유럽 순으로 표기한 후 영문 국가명에 따른 알파벳 순으로 정렬하였음

같은 표본을 대상으로 구한, 긍정적 정서 및 부정적 정서 척도의 원점수 평균에 대한 기술통계치는 <표 4-2>와 같다. 원점수 평균이란 각 구인을 측정하는 4점 척도 문항에 대한 응답의 평균을 의미한다. 긍정적 정서 척도와 부정적 정서 척도 모두에서 한국의 평균은 전체 평균보다 컸으며, 긍정적 정서의 평균이 가장 큰 국가는 멕시코(3.386), 부정적 정서의 평균이 가장 큰 국가는 일본(2.695)이었다. Kline(2005)이 제시한 기준에 따라, 모든 구인들의 왜도의 절댓값이 3보다 작고, 첨도의 절댓값이 10보다 작으므로, 자료는 정규분포를 따르고 있다고 볼 수 있다.

제 2 절 응답양식 분석

1. 모형 적합도

32개국 64,000명의 전체 9문항 응답에 대한 GRM, MNRM, IRTree, 간명 IRTree 모형의 적합도 지수는 <표 4-3>과 같다. 이때, 부정적 정서를 묻는 문항들은 역코딩하였다. 이 경우, MNRM이 가장 작은 AIC, BIC, SABIC 값을 가지는 것으로 나타났다. 정보지수 값이 작을수록 적합한 모형임을 의미하므로, MNRM이 가장 적합하고, GRM이 가장 부적합한 것으로 나타났다. 따라서 응답의 서열성을 가정하여 하나의 잠재변인의 수준만을 추정하는 GRM보다, ERS를 측정하고자 하는 잠재변인과 분리하여 해당 잠재변인의 수준을 추정하는 MNRM과 IRTree 모형이 자료를 분석하기에 더 적절하다고 판단되었다.

<표 4-3> 주관적 안녕감 척도에 대한 모형 적합도 지수

모형	AIC	BIC	SABIC
GRM	1132459	1132786	1132671
MNRM	1088150	1088712	1088515
IRTree	1102117	1102633	1102452
간명 IRTree	추정 불가	추정 불가	추정 불가

또한, 9문항이 주관적 안녕감을 측정하는 문항들이기는 하지만, PISA 2018에서 긍정적 정서와 부정적 정서를 나누어 해석하고 있으며(OECD, 2019), 긍정적 정서와 부정적 정서는 독립적인 차원이라는 점(Diener & Emmons, 1985; Warr et al., 1983; Watson & Clark, 1997)을 고려하여,

긍정적 정서를 묻는 5문항과 부정적 정서를 묻는 4문항을 나누어서 분석하고, 그 결과를 각각 <표 4-4>와 <표 4-5>에서 제시하였다.

<표 4-4> 긍정적 정서 척도에 대한 모형 적합도 지수

모형	AIC	BIC	SABIC
GRM	559549.5	559730.9	559667.3
MNRM	562553.8	562862.0	562754.0
IRTree	551042.6	551341.8	551237.0
간명 IRTree	추정 불가	추정 불가	추정 불가

<표 4-5> 부정적 정서 척도에 대한 모형 적합도 지수

모형	AIC	BIC	SABIC
GRM	534136.6	534281.6	534230.8
MNRM	532009.4	532254.2	532168.4
IRTree	521838.6	522083.4	521997.6
간명 IRTree	추정 불가	추정 불가	추정 불가

적합도 분석 결과, 두 개의 척도에서 모두 IRTree 모형이 가장 작은 AIC, BIC, SABIC 값을 가지는 것으로 나타났다. 따라서 응답양식이 측정하고자 하는 잠재변인의 수준을 보상할 수 있다고 가정하는 MNRM보다 응답양식과 측정하고자 하는 잠재변인이 독립적으로 기능한다고 가정하는 IRTree 모형이 자료를 분석하기에 더 적절하다고 판단되었다. 또한, 간명 IRTree 모형의 경우, 데이터에 적합하지 않아서 모수가 추정되지 않는다는 점에서, ERS의 방향성(긍정 또는 부정)이 존재한다고 볼 수 있다. 이는 간명 IRTree 모형에 비해 일반 IRTree 모형의 적합도가 더 높았다고 보고한 장윤선 외(2019)의 결과와 일치하며, 간명 IRTree 모형과 방향적 불변성이라는 동일한 전제를 가정하는 표준 IRTree 모형

보다 일반 IRTree 모형의 적합도가 더 높았다고 보고한 Jeon & De Boeck(2019)의 결과와도 같은 맥락으로 볼 수 있다.

이상의 결과를 종합해 볼 때, 주관적 안녕감 척도를 분석하기 위한 측정모형으로는 측정하고자 하는 구인 외 응답양식의 존재를 가정하는 모형이 더 적합하며, 정서에 대한 선행 연구를 바탕으로, 긍정적 정서 척도와 부정적 정서 척도를 나누어 분석한 경우에 더 적합한 모형인 IRTree 모형이 문항 분석에 가장 적절하다고 판단하였다.

2. 추정된 모수들의 분포

긍정적 정서 및 부정적 정서 척도 응답에서 한국과 한국을 포함한 OECD 전체 국가의 ERS 통제 전 주관적 안녕감에 대한 잠재변인 추정치(θ_g)와 ERS 통제 후 주관적 안녕감에 대한 잠재변인 추정치(θ_1), 부정적 ERS를 나타내는 잠재변수 추정치(θ_2), 긍정적 ERS를 나타내는 잠재변수 추정치(θ_3)의 최소값, 최대값, 평균, 표준편차, 사분위수는 아래 <표 4-6>, <표 4-7>과 같다. θ_g는 GRM을 통해 추정되었으며, θ_1, θ_2, θ_3는 IRTree 모형을 통해 추정되었다. 긍정적 정서 척도에 대한 θ_2를 제외하고, ERS 추정치의 평균과 중앙값 모두 한국 학생이 전체 학생의 ERS 추정치의 평균과 중앙값보다 컸다. 그리고 두 척도에서 한국 학생의 θ_g와 θ_1의 평균과 중앙값 모두 전체 학생의 θ_g와 θ_1의 평균과 중앙값보다 컸다. 그리고 두 척도에서 모두 θ_g보다 θ_1에서 표준편차가 감소하였다.

<표 4-6> 긍정적 정서 척도 응답 모수 추정치의 기술통계

	모수	최소값	최대값	평균	표준편차	사분위수		
						Q1	Q2	Q3
한국	θ_g	-3.042	1.515	.086	1.035	-.571	-.027	1.154
	θ_1	-2.443	1.079	.035	.894	-.544	.160	.895
	θ_2	-2.708	.503	-.016	.537	-.141	.159	.257
	θ_3	-1.621	1.479	.086	.978	-.810	-.013	1.145
전체	θ_g	-3.042	1.515	.000	.911	-.571	-.096	.740
	θ_1	-2.443	1.079	.000	.799	-.424	.043	.625
	θ_2	-2.708	.503	.001	.551	-.155	.183	.316
	θ_3	-1.621	1.479	-.002	.868	-.737	-.101	.819

θ_g: ERS 통제 전 주관적 안녕감에 대한 잠재변인 추정치

θ_1: ERS 통제 후 주관적 안녕감에 대한 잠재변인 추정치

θ_2: 부정적 ERS 추정치, θ_3: 긍정적 ERS 추정치

<표 4-7> 부정적 정서 척도 응답 모수 추정치의 기술통계

	모수	최소값	최대값	평균	표준편차	사분위수		
						Q1	Q2	Q3
한국	θ_g	-2.147	2.550	.105	1.021	-.535	.168	.906
	θ_1	-1.382	1.694	.139	.872	-.507	.112	.785
	θ_2	-.756	1.994	.110	.719	-.460	-.144	.484
	θ_3	-.751	2.321	.222	.716	-.246	.029	.651
전체	θ_g	-2.147	2.550	.000	.880	-.535	-.052	.595
	θ_1	-1.382	1.694	-.001	.834	-.887	-.017	.693
	θ_2	-.756	1.994	-.001	.691	-.474	-.239	.418
	θ_3	-.751	2.321	-.001	.642	-.410	-.127	.114

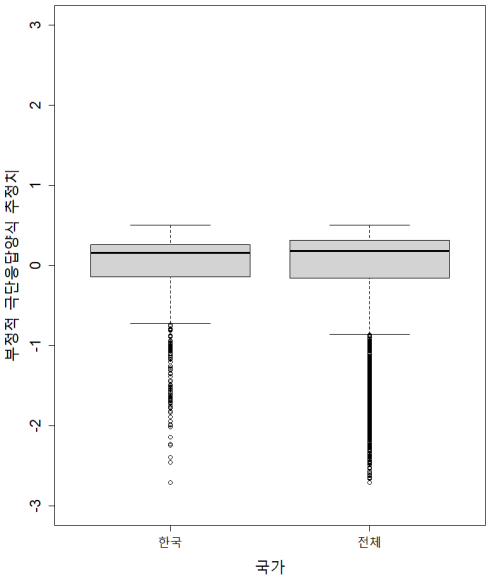
θ_g: ERS 통제 전 주관적 안녕감에 대한 잠재변인 추정치

θ_1: ERS 통제 후 주관적 안녕감에 대한 잠재변인 추정치

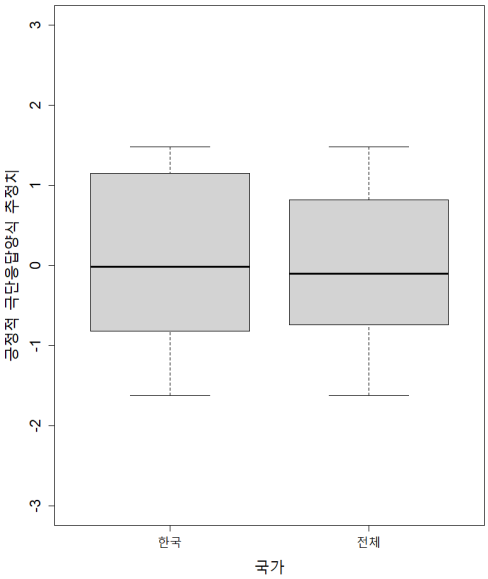
θ_2: 부정적 ERS 추정치, θ_3: 긍정적 ERS 추정치

[그림 4-1], [그림 4-2], [그림 4-3], [그림 4-4]에서 한국과 OECD 전체 국가의 ERS를 척도 및 방향별로 비교하였다. [그림4-1]은 긍정적 정서 척도에 대한 한국과 전체 국가의 θ_2를 상자도표로 나타낸 것이다. 한국과 전체 국가의 θ_2의 분포는 유사하지만, 나머지 ERS 추정치들과 달리, 전체 국가의 θ_2의 중앙값과 3사분위수가 한국의 θ_2의 중앙값과 3사분위수보다 더 큰 것을 알 수 있다. [그림 4-2]는 긍정적 정서 척도에 대한 한국과 전체 국가의 θ_3를 상자도표로 나타낸 것이다. 긍정적 정서 척도에 대한 θ_3의 경우, 한국이 전체 국가보다 중앙값과 3사분위수가 더 크다. 또한, 다른 ERS 추정치들보다 θ_3의 범위가 크다는 점에서 긍정적 정서 척도에 대한 θ_3의 경우, 개인 간 편차가 크다는 것을 알 수 있다.

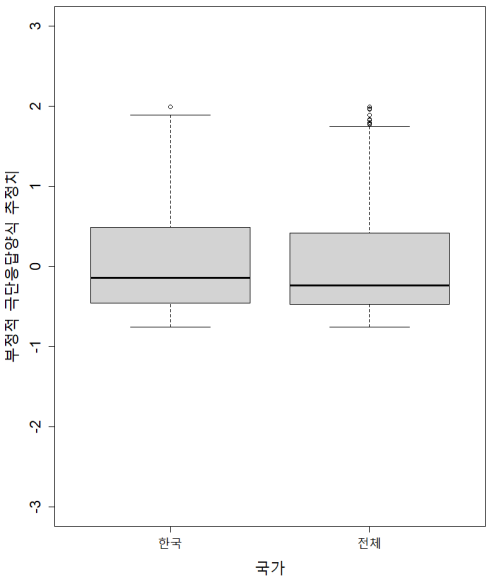
[그림 4-3]은 부정적 정서 척도에 대한 한국과 전체 국가 간 θ_2를 비교한 상자도표이다. 이 경우, 한국과 전체 국가의 θ_2의 분포가 유사하지만, 한국의 θ_2의 중앙값과 3사분위수가 전체 국가의 θ_2의 중앙값과 3사분위수보다 크다는 것을 알 수 있다. [그림 4-4]는 부정적 정서 척도에 대한 한국과 전체 국가 간 θ_3를 비교한 상자도표이다. 한국이 전체 국가보다 부정적 정서 척도에 대한 θ_3의 개인 간 편차가 큰 것을 확인할 수 있다. 또한, 앞선 [그림 4-2] 및 [그림 4-3]과 마찬가지로, 한국의 θ_3의 중앙값과 3사분위수가 전체 국가의 θ_3의 중앙값과 3사분위수보다 큰 것을 확인할 수 있다.



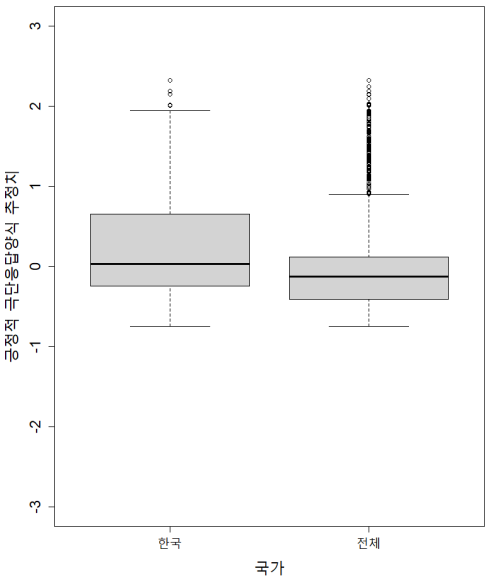
[그림 4-1] 긍정적 정서 척도의 부정적 ERS 추정치에 대한 상자도표



[그림 4-2] 긍정적 정서 척도의 긍정적 ERS 추정치에 대한 상자도표



[그림 4-3] 부정적 정서 척도의 부정적 ERS 추정치에 대한 상자도표



[그림 4-4] 부정적 정서 척도의 긍정적 ERS 추정치에 대한 상자도표

3. 추정된 모수들의 상관관계

ERS를 통제하지 않은 주관적 안녕감에 대한 잠재변인, ERS를 통제한 주관적 안녕감에 대한 잠재변인, ERS에 대한 잠재변인들 간 관계를 확인하고자 전체 학생의 응답을 대상으로 추정된 모수들 간 Pearson의 상관계수를 산출하고, 이를 <표 4-8>에 제시하였다.

긍정적 정서와 부정적 정서 척도에서 산출된 θ_g와 θ_1의 상관 관계는 .933, .949로 GRM과 IRTree 모형에서 산출된 주관적 안녕감에 대한 잠재변인 간 상관이 매우 높다는 것을 알 수 있다. 그리고 두 척도의 θ_g 간 상관계수는 -.299, 두 척도의 θ_1 간 상관계수는 -.284로, ERS 통제 전후 두 척도에서 주관적 안녕감에 대한 잠재변인 추정치 간 상관관계는 변함이 없음을 알 수 있다.

긍정적 정서 척도에서 θ_3와 θ_1 간 상관계수는 .807, θ_2와 θ_1 간 상관계수는 .688, 부정적 정서 척도에서 θ_3와 θ_1 간 상관계수는 .654로, 모두 높은 정적인 상관을 보였다. 반면, 부정적 정서 척도에 대한 θ_2와 θ_1 간 상관계수는 -.588로, 높은 부적 상관을 보였다. 또한, θ_2와 θ_3 간 상관계수들은 -.340에서 .165 사이의 값을 가지며, 상관이 낮은 것으로 나타났다. 한국 학생을 대상으로 한 GRM과 IRTree 모형에서 산출된 모수들 간 상관관계는 <표 4-9>에 제시되어 있다. <표 4-9>의 상관관계는 앞서 서술한 전체 학생의 응답을 대상으로 추정된 모수들 간 상관관계와 같은 양상을 나타냈다.

<표 4-8> 전체 학생의 응답 모수 추정치 간 상관관계

구분		긍정적 정서				부정적 정서			
		θ_g	θ_1	θ_2	θ_3	θ_g	θ_1	θ_2	θ_3
긍정적 정서	θ_g	-							
	θ_1	.933	-						
	θ_2	.463	.688	-					
	θ_3	.935	.807	.165	-				
부정적 정서	θ_g	-.299	-.292	-.104	-.305	-			
	θ_1	-.287	-.284	-.106	-.294	.949	-		
	θ_2	.106	.043	-.186	.200	-.669	-.588	-	
	θ_3	-.288	-.346	-.340	-.196	.630	.654	.105	-

θ_g: ERS 통제 전 주관적 안녕감에 대한 잠재변인 추정치

θ_1: ERS 통제 후 주관적 안녕감에 대한 잠재변인 추정치

θ_2: 부정적 ERS 추정치, θ_3: 긍정적 ERS 추정치

<표 4-9> 한국 학생의 응답 모수 추정치 간 상관관계

구분		긍정적 정서				부정적 정서			
		θ_g	θ_1	θ_2	θ_3	θ_g	θ_1	θ_2	θ_3
긍정적 정서	θ_g	-							
	θ_1	.953	-						
	θ_2	.525	.705	-					
	θ_3	.957	.865	.287	-				
부정적 정서	θ_g	-.288	-.301	-.234	-.255	-			
	θ_1	-.278	-.291	-.221	-.249	.961	-		
	θ_2	.244	.203	.017	.270	-.697	-.661	-	
	θ_3	-.168	-.228	-.312	-.098	.690	.674	-.013	-

θ_g: ERS 통제 전 주관적 안녕감에 대한 잠재변인 추정치

θ_1: ERS 통제 후 주관적 안녕감에 대한 잠재변인 추정치

θ_2: 부정적 ERS 추정치, θ_3: 긍정적 ERS 추정치

제 3 절 국가별 극단응답양식 추정치 비교

1. 한국과 그 외 국가들의 극단응답양식 추정치 비교

한국과 한국을 제외한 국가들 간 ERS의 차이가 있는지 알아보기 위하여 맨-휘트니 U 검정을 사용하였다. 샤피로-윌크 검정으로 한국의 ERS 추정치에 대해 정규성을 검정한 결과, 정규성을 만족하지 않는 것으로 나타났다. 또한, 앤더슨-달링 검정으로, 한국을 제외한 국가 전체의 ERS 추정치에 대해 정규성을 검정한 결과, 마찬가지로 정규성을 만족하지 않는 것으로 나타났다. 이때, 정규성 검정을 위해서 샤피로-윌크 검정이 아닌, 앤더슨-달링 검정을 사용한 이유는 샤피로-윌크 검정은 5,000개 이하의 표본 크기를 가진 표본에서만 사용할 수 있기 때문이다. 두 집단의 ERS 추정치는 정규분포를 따르지 않으므로, 특정한 분포를 가정하지 않고 두 표본 집단이 같은 모집단에서 추출되었는지를 검정하는 맨-휘트니 U 검정이 적절하다고 판단되었다(McKnight & Najab, 2010).

맨-휘트니 U 검정 결과는 <표 4-10>에 제시되어 있다. <표 4-10>에서 평균 차이와 중앙값 차이는 한국의 ERS 추정치에 대한 평균 및 중앙값에서 그 외 국가의 ERS 추정치에 대한 평균 및 중앙값을 뺀 값이다. 모든 ERS에서 두 집단 간 차이가 유의했으며, 한국 학생들이 그 외 국가들의 학생들보다 긍정적 정서에 대한 θ_3, 부정적 정서에 대한 θ_3 및 θ_2의 크기가 크다고 할 수 있다. 반면, 긍정적 정서에 대한 θ_2의 경우, 한국의 추정치가 한국을 제외한 국가들에 비해 작았다.

<표 4-10> 한국과 그 외 국가 간 ERS 추정치 차이에 대한 맨-휘트니 U 검정 결과

	모수	국가	중앙값	중앙값 차이	W	유의 확률
긍정적 정서	θ_2	한국	.159	-.024	66348881	<.001
		그 외	.183			
	θ_3	한국	-.013	.088	59205241	<.001
		그 외	-.101			
부정적 정서	θ_2	한국	-.144	.095	56316725	<.001
		그 외	-.239			
	θ_3	한국	.029	.156	49139632	<.001
		그 외	-.127			

θ_2: 부정적 ERS 추정치, θ_3: 긍정적 ERS 추정치

2. 국가별 극단응답양식 추정치 비교

국가별 ERS 추정치 또한 정규성을 만족하지 않기 때문에, 국가별 ERS의 차이를 검정하고자, 크루스칼-왈리스 검정을 실시하였다. 검정 결과, <표 4-11>에 제시되어 있듯이, 모든 국가에서 척도별 ERS 추정치의 중앙값이 같지는 않은 것으로 나타났다.

<표 4-11> OECD 국가별 극단응답양식 추정치 차이에 대한 크루스칼-왈리스 검정 결과

	모수	χ^2	자유도	유의 확률
긍정적 정서	θ_2	2089.6	31	<.001
	θ_3	2059.9	31	<.001
부정적 정서	θ_2	1470.3	31	<.001
	θ_3	1597.2	31	<.001

θ_2: 부정적 ERS 추정치, θ_3: 긍정적 ERS 추정치

이에 따라, Hochberg와 Benjamini-Hochberg를 사용하여 사후검정을 수행하였다. 비모수 통계분석 이후의 사후분석으로는 주로 Bonferroni 방법이 사용된다. 그러나 본 연구와 같이 시행되는 검정의 수가 많은 경우에 Bonferroni 방법을 사용하면 제 2종의 오류가 지나치게 증가한다는 단점이 있다. 이에 따라, Bonferroni 방법보다 덜 보수적인, 단계적 유의수준 보정법인 Hochberg 방법을 사용하였다(Lee & Lee, 2018). 또한, Bonferroni 및 Hochberg와 같은 FWER을 통제하는 방법들과 달리 FDR을 통제하는 방법인 Benjamini-Hochberg도 사용하여, 그 결과를 비교하였다.

사후 검정 결과는 아래 <표 4-12>와 <표 4-13>에 제시되어 있다. 표에서 중앙값의 차이는 한국의 ERS 추정치의 중앙값에서 다른 국가의 ERS 추정치의 중앙값을 뺀 값이다. 그리고 Hochberg 방법과 Benjamini-Hochberg 방법을 통해 구한 사후검정 결과의 유의확률 값을 각각 Hochberg 유의확률과 BH 유의확률 열에 제시하였다. Hochberg 방법과 Benjamini-Hochberg 방법 간 통계적 유의성에서 차이가 있는 사례는 각 ERS 별로 3-4 개로, 사후검정 방법에 따른 차이는 크지 않은 것으로 나타났다. 보다 엄격하게 결과를 해석하기 위해 두 방법에서 모두 유의한 결과만을 통계적으로 유의한 것으로 간주하였다.

긍정적 정서 척도에 대한 θ_2의 경우, 앞선 연구 결과와 마찬가지로, 대다수의 국가들에 비해 한국의 θ_2의 크기가 작은 것으로 나타났다. 구체적으로, 호주, 캐나다, 멕시코, 미국, 스위스, 체코, 독일, 덴마크, 스페인, 핀란드, 영국, 그리스, 아일랜드, 아이슬란드, 이탈리아, 리투아니아, 룩셈부르크, 네덜란드, 폴란드, 포르투갈, 스웨덴에 비해 한국의 긍정적 정서 척도에 대한 θ_2의 크기가 작았다. 반면, 이탈리아와 슬로베니아에 비해서는 한국의 긍정적 정서 척도에 대한 θ_2의 크기가 컸다.

긍정적 정서 척도에 대한 θ_3의 경우, 일본, 캐나다, 체코, 에스토니아, 핀란드, 영국, 아일랜드, 아이슬란드, 이탈리아, 라트비아, 폴란드, 슬로바키아, 슬로베니아, 스웨덴, 터키에 비해 한국의 θ_3의 크기가 컸다. 그리고 멕시코와 콜롬비아, 스페인에 비해 한국의 θ_3의 크기가 작았다.

부정적 정서 척도에 대한 θ_2의 경우, 일본, 호주, 캐나다, 미국, 콜롬비아, 스위스, 체코, 독일, 스페인, 에스토니아, 핀란드, 영국, 그리스, 아일랜드, 이탈리아, 룩셈부르크, 라트비아, 네덜란드, 폴란드, 포르투갈, 슬로바키아, 스웨덴에 비해 한국의 θ_2의 크기가 컸다. 반면, 덴마크와 터키에 비해 한국의 θ_2의 크기가 작았다. 그리고 부정적 정서 척도에 대한 θ_3의 경우, 일본을 제외한 모든 국가에 비해 한국의 θ_3의 크기가 컸다.

이상의 결과를 종합해보면, 앞선 한국과 한국을 제외한 국가 간 맨-휘트니 U 검정 결과와 일맥상통하게, 대다수의 국가에 비해 한국 학생들의 긍정적 정서 척도에 대한 θ_3, 부정적 정서 척도에 대한 θ_3 및 θ_2가 유의하게 컸다.

<표 4-12> OECD 국가별 긍정적 정서 척도의
ERS 추정치 차이에 대한 사후검정 결과

국가	θ_2			θ_3		
	중앙값 차이	Hochberg 유의확률	BH 유의확률	중앙값 차이	Hochberg 유의확률	BH 유의확률
일본	.000	.974	.015	.149	<.001	<.001
호주	-.024	<.001	<.001	-.047	.972	.072
캐나다	-.038	<.001	<.001	.134	<.001	<.001
멕시코	-.038	<.001	<.001	-.328	<.001	<.001
미국	-.024	<.001	<.001	.163	<.001	<.001
칠레	-.024	.261	.002	-.132	.972	.059
콜롬비아	-.024	.083	<.001	-.146	.003	<.001
스위스	-.038	<.001	<.001	-.077	.972	.136
체코	-.024	.032	<.001	.262	<.001	<.001
독일	-.038	<.001	<.001	.019	.972	.054
덴마크	-.056	<.001	<.001	-.267	.374	.004
스페인	-.038	<.001	<.001	-.132	.006	<.001
에스토니아	.000	.420	.004	.114	<.001	<.001
핀란드	-.056	<.001	<.001	.725	<.001	<.001
프랑스	.000	.076	<.001	-.143	.972	.040
영국	-.043	<.001	<.001	.673	<.001	<.001
그리스	-.043	<.001	<.001	.002	.972	.241
헝가리	-.024	.974	.670	-.286	.489	.005
아일랜드	-.055	<.001	<.001	.104	<.001	<.001
아이슬란드	-.043	<.001	<.001	.418	.005	<.001
이탈리아	.354	<.001	<.001	.035	<.001	<.001
리투아니아	-.024	<.001	<.001	.018	.972	.679
룩셈부르크	-.024	<.001	<.001	.019	.972	.727
라트비아	.000	.009	<.001	.049	.010	<.001
네덜란드	-.055	<.001	<.001	-.030	.972	.695
폴란드	-.024	<.001	<.001	.171	<.001	<.001
포르투갈	-.043	<.001	<.001	-.132	.972	.177
슬로바키아	.000	.135	.001	.114	<.001	<.001
슬로베니아	.104	.001	<.001	.725	<.001	<.001
스웨덴	-.043	<.001	<.001	.114	<.001	<.001
터키	.000	.974	.207	.243	<.001	<.001

θ_2: 부정적 ERS 추정치, θ_3: 긍정적 ERS 추정치, BH: Benjamini-Hochberg

<표 4-13> OECD 국가별 부정적 정서 척도의
ERS 추정치 차이에 대한 사후검정 결과

국가	θ_2			θ_3		
	중앙값 차이	Hochberg 유의확률	BH 유의확률	중앙값 차이	Hochberg 유의확률	BH 유의확률
일본	.180	<.001	<.001	-.085	<.001	<.001
호주	.157	<.001	<.001	.243	<.001	<.001
캐나다	.095	.017	<.001	.112	<.001	<.001
멕시코	.095	.992	.531	.156	<.001	<.001
미국	.095	<.001	<.001	.112	<.001	<.001
칠레	.095	.992	.011	.156	<.001	<.001
콜롬비아	.095	.013	<.001	.148	<.001	<.001
스위스	.157	<.001	<.001	.171	<.001	<.001
체코	.180	<.001	<.001	.112	<.001	<.001
독일	.179	<.001	<.001	.212	<.001	<.001
덴마크	-.517	<.001	<.001	.201	<.001	<.001
스페인	.180	<.001	<.001	.112	<.001	<.001
에스토니아	.157	<.001	<.001	.148	<.001	<.001
핀란드	.159	<.001	<.001	.275	<.001	<.001
프랑스	.095	.170	.001	.148	<.001	<.001
영국	.180	<.001	<.001	.067	<.001	<.001
그리스	.095	<.001	<.001	.112	<.001	<.001
헝가리	.095	.992	.077	.171	<.001	<.001
아일랜드	.162	<.001	<.001	.156	<.001	<.001
아이슬란드	.095	.992	.245	.156	<.001	<.001
이탈리아	.157	<.001	<.001	.156	<.001	<.001
리투아니아	.095	.992	.020	.197	<.001	<.001
룩셈부르크	.162	<.001	<.001	.144	<.001	<.001
라트비아	.098	<.001	<.001	.148	<.001	<.001
네덜란드	.157	<.001	<.001	.229	<.001	<.001
폴란드	.095	.015	<.001	.192	<.001	<.001
포르투갈	.180	<.001	<.001	.112	<.001	<.001
슬로바키아	.157	<.001	<.001	.144	<.001	<.001
슬로베니아	.095	.058	<.001	.302	<.001	<.001
스웨덴	.157	<.001	<.001	.156	<.001	<.001
터키	-.463	<.001	<.001	.083	<.001	<.001

θ_2: 부정적 ERS 추정치, θ_3: 긍정적 ERS 추정치, BH: Benjamini-Hochberg

제 4 절 극단응답양식 통제 후 주관적 안녕감에 대한 잠재변인 추정치의 변화

1. 한국의 극단응답양식 통제 후 주관적 안녕감에 대한 잠재변인 추정치의 변화

한국 학생들에게서 ERS 통제 후 주관적 안녕감에 대한 잠재변인 수준의 유의한 변화가 있었는지를 알아보기 위해, 윌콕슨 부호 순위 검정를 사용하였다. 윌콕슨 부호 순위 검정은 관측값이 쌍을 이루는 경우, 두 관측값의 차이의 순위합으로 두 표본의 차이를 검정하는 비모수적 검정 방법으로, 한국 학생들의 θ_g및 θ_1처럼 정규성을 만족하지 않는 두 종속 표본 간 차이 검정에 적합하다.

<표 4-14> ERS 통제 전후 한국의 주관적 안녕감에 대한 잠재변인 추정치 차이에 대한 윌콕슨 부호 순위 검정 결과

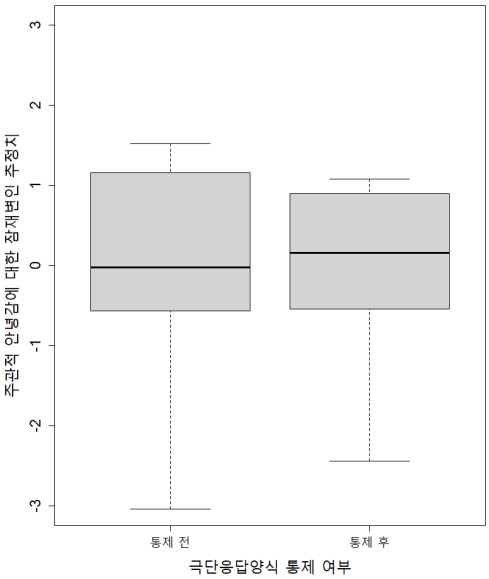
	모수	중앙값	중앙값 차이	V	유의 확률
긍정적 정서	θ_g	-.027	.187	1127551	<.001
	θ_1	.160			
부정적 정서	θ_g	.168	-.056	825714	<.001
	θ_1	.112			

θ_g: ERS 통제 전 주관적 안녕감에 대한 잠재변인 추정치

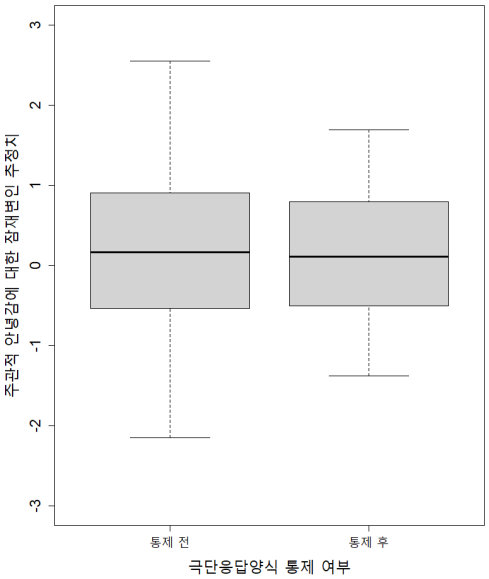
θ_1: ERS 통제 후 주관적 안녕감에 대한 잠재변인 추정치

<표 4-14>는 한국 학생의 θ_g와 θ_1의 차이에 대한 윌콕슨 부호 순위 검정 결과를 제시하고 있다. <표 4-14>에서 중앙값 차이는 θ_1의 중앙값

에서 θ_g의 중앙값을 뺀 값이다. 검정 결과, 긍정적 정서 척도에서는 θ_g에 비해 θ_1이 유의하게 증가했으나, 부정적 정서 척도에서는 θ_g에 비해 θ_1이 유의하게 감소하였다. 즉, ERS 통제 후, 한국 학생의 긍정적 정서에 대한 잠재변인 수준은 증가했으나, 부정적 정서에 대한 잠재변인 수준은 감소했다. [그림 4-5]와 [그림 4-6]은 각 척도에서 한국 학생의 θ_g와 θ_1을 비교하는 상자도표이다. [그림 4-5]에서 ERS 통제 후 긍정적 정서 척도에서 한국의 주관적 안녕감에 대한 잠재변인 추정치가 커진 것을 확인할 수 있으며, [그림 4-6]에서는 ERS 통제 후 부정적 정서 척도에서 한국의 주관적 안녕감에 잠재변인 추정치가 작아진 것을 확인할 수 있다.



[그림 4-5] 긍정적 정서 척도에서 ERS 통제 전후 한국의 주관적 안녕감에 대한 잠재변인 추정치 상자도표



[그림 4-6] 부정적 정서 척도에서 ERS 통제 전후 한국의 주관적 안녕감에 대한 잠재변인 추정치 상자도표

2. 한국과 그 외 국가들의 주관적 안녕감에 대한 잠재변인 추정치 비교

θ_g와 θ_1에 대하여 한국과 그 외 국가 간 유의한 차이가 있는지를 알아보기 위하여 맨-휘트니 U 검정을 실시하였다. 샤피로-윌크 검정으로 한국의 θ_g와 θ_1에 대해 정규성을 검정한 결과, 정규성을 만족하지 않는 것으로 나타났다. 또한, 앤더슨-달링 검정을 통해 한국을 제외한 전체 국가들의 θ_g와 θ_1에 대해 정규성을 검정한 결과, 마찬가지로 정규성을 만족하지 않는 것으로 나타났다. 두 집단의 θ_g와 θ_1은 정규분포를 따르지 않으므로, 특정한 분포를 가정하지 않고 두 표본 집단이 같은 모집단에서 추출되었는지를 검정하는 맨-휘트니 U 검정이 적절하다고 판단되었다(McKnight & Najab, 2010).

<표 4-15> 한국과 그 외 국가 간 주관적 안녕감에 대한 잠재변인 추정치 차이에 대한 맨-휘트니 U 검정 결과

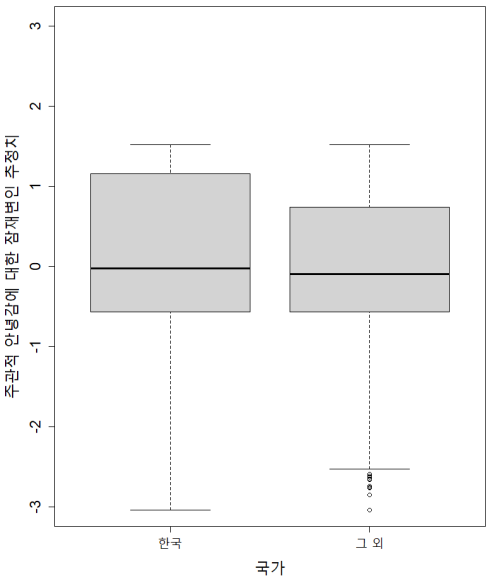
	모수	국가	중앙값	중앙값 차이	W	유의확률
긍정적 정서	θ_g	한국	-.027	.069	58783880	<.001
		그 외	-.096			
	θ_1	한국	.160	.117	58647998	<.001
		그 외	.043			
부정적 정서	θ_g	한국	.168	.233	58009301	<.001
		그 외	-.065			
	θ_1	한국	.112	.129	56206140	<.001
		그 외	-.017			

θ_g: ERS 통제 전 주관적 안녕감에 대한 잠재변인 추정치

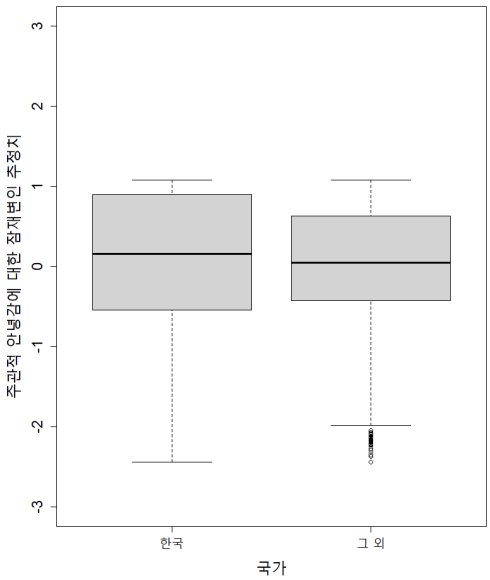
θ_1: ERS 통제 후 주관적 안녕감에 대한 잠재변인 추정치

맨-휘트니 U 검정 결과는 <표 4-15>에 제시되어 있다. <표 4-15>에서 중앙값 차이는 한국의 주관적 안녕감에 대한 잠재변인 추정치에서 그 외 국가의 주관적 안녕감에 대한 잠재변인 추정치를 뺀 값이다. 두 척도 모두에서 ERS 통제 전과 후에 한국과 그 외 국가 간 θ_g와 θ_1에 대한 차이가 유의했다. 그리고 ERS 통제 여부와 관계 없이, 한국 학생들이 그 외 국가의 학생들보다 두 척도에서 모두 θ_g와 θ_1의 크기가 큰 것을 알 수 있다. 즉, ERS 통제 전과 후에 한국이 그 외 국가에 비해 긍정적 정서 및 부정적 정서에 대한 잠재변인 수준이 모두 높다. 그리고 긍정적 정서 척도에서는 ERS 통제 후 한국과 그 외 국가 간 주관적 안녕감에 대한 잠재변인 추정치의 차이가 커진 반면, 부정적 정서 척도에서는 ERS 통제 후 한국과 그 외 국가 간 주관적 안녕감에 대한 잠재변인 추정치의 차이가 줄어든 것을 확인할 수 있다.

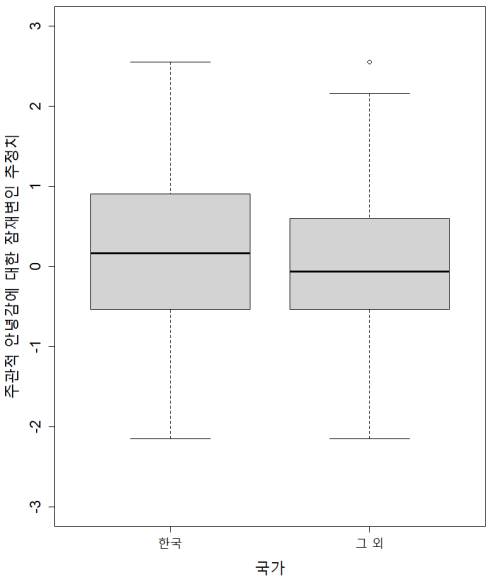
[그림 4-7]과 [그림 4-8]은 긍정적 정서 척도에서 한국과 그 외 국가의 θ_g와 θ_1을 비교하는 상자도표이다. 한국의 θ_g와 θ_1의 중앙값이 그 외 국가의 θ_g와 θ_1의 중앙값보다 크다는 것을 확인할 수 있으며, 한국과 그 외 국가 간 중앙값의 차이는 θ_g보다 θ_1에서 더 크다. 또한, [그림 4-9], [그림 4-10]은 부정적 정서 척도에서 한국과 그 외 국가의 θ_g와 θ_1을 비교하는 상자도표이다. 마찬가지로, 한국의 θ_g와 θ_1의 중앙값이 그 외 국가의 θ_g와 θ_1의 중앙값보다 크다는 것을 확인할 수 있다. 하지만 한국과 그 외 국가 간 중앙값의 차이는 θ_g보다 θ_1에서 더 작다.



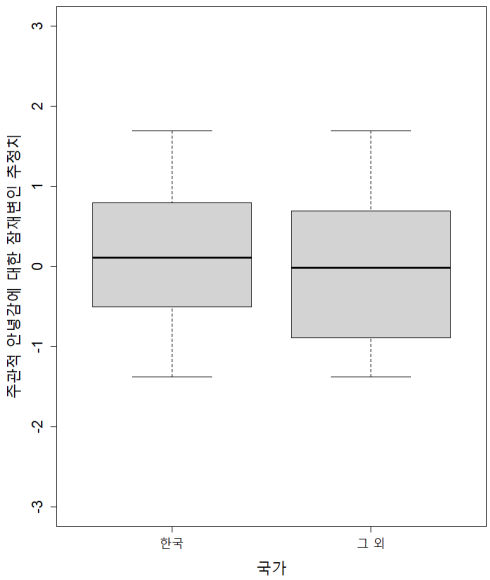
[그림 4-7] 긍정적 정서 척도에서 ERS 통제 전 주관적 안녕감에 대한 잠재변인 추정치 상자도표



[그림 4-8] 긍정적 정서 척도에서 ERS 통제 후 주관적 안녕감에 대한 잠재변인 추정치 상자도표



[그림 4-9] 부정적 정서 척도에서 ERS 통제 전 주관적 안녕감에 대한 잠재변인 추정치 상자도표



[그림 4-10] 부정적 정서 척도에서 ERS 통제 후 주관적 안녕감에 대한 잠재변인 추정치 상자도표

제 5 장 논의 및 결론

제 1 절 요약 및 논의

PISA 2018 조사 결과, 10점 척도로 측정된 우리나라 학생들의 삶에 대한 만족도의 평균은 OECD 평균보다 낮으나, 4점 척도로 측정된 주관적 안녕감 척도를 통해 산출된 긍정적 감정 지수의 평균은 OECD 평균보다 높은 것으로 나타났다. PISA 2018과 같은 자기보고식 검사는 심리적 구인을 측정하는 데 있어 제작하기 간편하고, 실시가 용이하며, 단시간 내 많은 자료를 수집할 수 있다는 면에서 많은 사회과학 분야에서 선호되고 있지만, 해당 구인을 간접적으로 측정한다는 점에서 측정의 오차를 항상 동반한다(손원숙, 2015).

전형적인 응답 편파의 유형 중 하나인 ERS는 응답 척도의 가장 높거나 가장 낮은 범주를 지지하는 경향(Ames & Leventhal, 2021)이다. 청소년의 안녕감에 관한 연구들에 따르면, 여러 응답양식 중, ERS로 인한 응답의 차이가 가장 많이 보고 되었으며, ERS는 국가의 영향을 받았다(Chen et al., 1995; Clefberg et al., 2015; Gilman et al., 2008; Isabel et al., 2016; Lee et al., 2020).

선행연구들에 따르면, 응답범주 수가 작을수록, ERS의 수준이 증가하고(Weijters et al., 2010a), 이로 인해 측정하고자 하는 구인과 무관하게, 서로 다른 문화적 집단 간 응답 결과의 차이가 발생했다(Hui & Triandis, 1989). 따라서, 삶의 만족도에 비해 긍정적 감정 지수가 높은 이유가 주관적 안녕감 척도 응답 시 ERS가 발생했기 때문인지를 본 연

구를 통해 알아보고자 했다.

여러 문항반응모형들 중 최근 응답양식 연구에서는 IRTree 모형과 MNRM이 가장 활발하게 사용되고 있다. 이는 두 모형이 응답의 비서열성을 가정할 수 있기 때문이며, 특히나 IRTree 모형은 응답자의 의사결정 과정에 대한 가정을 바탕으로, 비교적 자유롭게 모형을 구현할 수 있다는 장점을 가진다(장윤선 외, 2019).

이에 따라, OECD 국가별 2,000명씩 무선표집한 주관적 안녕감 척도 응답 데이터를 분석하기 위해 IRTree 모형이 적합한지 검증하고자, MNRM, IRTree 모형, 간명 IRTree 모형, GRM의 적합도를 비교하였다. 또한, ERS 통제 후 추정된 모수의 신뢰성을 파악하고자, ERS 통제 전후 추정된 모수들의 상관관계를 분석하였다. 그리고 한국 학생들이 실제로 한국을 제외한 OECD 국가들의 학생들보다 극단응답을 하는 경향이 큰지를 비교하기 위해, 맨-휘트니 U 검정, 크루스칼-왈리스 검정과 그에 따른 사후비교를 활용하여, 한국과 한국을 제외한 OECD 국가들 간 ERS의 수준을 비교하였다. 다음으로, 한국 학생들의 ERS가 주관적 안녕감 추정에 유의한 영향을 미쳤는지를 파악하기 위해, 윌콕슨 부호 순위 검정을 사용하여 ERS 통제 전후 주관적 안녕감에 대한 잠재변인 추정치의 크기를 비교하였다. 끝으로, ERS 통제 전후 한국과 한국을 제외한 OECD 국가 간 주관적 안녕감에 대한 잠재변인 추정치의 차이가 유의한지 검증하고자, 맨-휘트니 U 검정을 사용하였고, 그 차이가 어떻게 변화하였는지를 비교하였다.

본 연구의 주요 연구 결과와 그에 따른 시사점은 다음과 같다.

첫째, 한국 학생의 PISA 2018 주관적 안녕감 척도 응답에는 ERS가 존재했고, 이를 분석하기에 가장 적합한 모형은 IRTree 모형이었다. GRM, MNRM, IRTree 모형, 간명 IRTree 모형으로 긍정적 정서 척도

및 부정적 정서 척도 응답을 분석한 결과, IRTree 모형이 가장 작은 AIC, BIC, SABIC 값을 가지는 것으로 나타났다. 따라서 응답양식이 측정하고자 하는 잠재변인의 수준을 보상할 수 있다고 가정하는 MNRM 보다 응답양식과 측정하고자 하는 잠재변인이 독립적으로 기능한다고 가정하는 IRTree 모형이 자료를 분석하기에 더 적절하다고 판단되었다. 또한, ERS의 방향성이 존재하지 않는다고 가정하는 간명 IRTree 모형의 경우, 데이터에 적합하지 않아 추정이 되지 않는다는 점에서, 주관적 안녕감에 대한 문항 응답에서 ERS의 방향성(긍정 또는 부정)이 존재한다고 볼 수 있다. 이는 간명 IRTree 모형에 비해 일반 IRTree 모형의 적합도가 높다고 보고한 장윤선 외(2019)의 연구 결과와 일치하며, 간명 IRTree 모형과 방향적 불변성이라는 동일한 전제를 가정하는 표준 IRTree 모형보다 일반화 IRTree 모형의 적합도가 더 높았다고 보고한 Jeon & De Boeck(2019)의 결과와도 같은 맥락으로 볼 수 있다.

둘째, GRM과 IRTree 모형을 통해 추정된 모수 간 상관분석을 통해 IRTree 모형의 신뢰성을 실증하였다. ERS를 통제하기 전의 주관적 안녕감에 대한 잠재변인 추정치(θ_g)와 ERS를 통제한 후의 주관적 안녕감에 대한 잠재변인 추정치(θ_1) 간에는 .933~.949의 높은 정적인 상관관계가 존재하였다. 또한, 두 척도의 주관적 안녕감에 대한 잠재변인 추정치 간 상관계수는 ERS 통제 전후에 일정하게 유지되었다. 이는 장윤선 외(2019)의 연구 결과와 상통하며, IRTree 모형에서 얻어진 θ_1이 리커트 척도의 응답 분석에서 사용되던 기존의 모형인 GRM에서 추정된 θ_g와 비교하여 크게 다르지 않음을 보였다고 할 수 있다. 아울러, 긍정적 정서 척도에 대한 긍정적 ERS 추정치(θ_3)와 부정적 ERS 추정치(θ_2), 부정적 정서 척도에 대한 θ_3의 경우, θ_1과의 상관이 .654~.807로 높은 정적인 상

관을 보였다. 이는 IRTree 모형을 활용하여 ERS를 분석한 선행연구들의 결과와 일치한다(Böckenholt, 2019; LaHuis et. al., 2018) 반면, 부정적 정서 척도에 대한 θ_2의 경우, θ_1과 -.588의 높은 부적 상관을 보였다.

셋째, 한국 학생들이 한국을 제외한 OECD 국가들에 비해 극단응답을 하는 경향이 더 크다는 것을 확인했다. 맨-휘트니 U 검정 및 크루스칼-왈리스 검정과 그에 따른 사후검정 결과, 한국 학생들의 긍정적 정서에 대한 θ_3, 부정적 정서에 대한 θ_3 및 θ_2의 크기가 그 외 국가들의 학생들보다 컸다. 반면, 한국의 긍정적 정서에 대한 θ_2의 경우, 한국을 제외한 국가들의 θ_2에 비해 작았다.

넷째, ERS가 한국 학생들의 주관적 안녕감에 대한 잠재변인에 유의한 영향을 미친다는 것을 확인했다. 한국의 θ_g와 θ_1에 대한 윌콕슨 부호 순위 검정 결과, 긍정적 정서 척도에서는 θ_g에 비해 θ_1이 유의하게 컸으나, 부정적 정서 척도에서는 θ_g에 비해 θ_1이 유의하게 작았다. 즉, ERS 통제 여부에 따라, 한국의 주관적 안녕감에 대한 잠재변인 수준이 달라졌다.

다섯째, ERS가 부정적 정서 척도에서 한국 학생들의 주관적 안녕감에 대한 잠재변인 수준이 한국을 제외한 다른 국가 학생들의 주관적 안녕감에 대한 잠재변인 수준에 비해 더 큰 것에 기여한다고 추론할 수 있다. 맨-휘트니 U 검정 결과에 따르면, 두 척도에서 모두 한국 학생들의 θ_g와 θ_1이 그 외 국가의 θ_g와 θ_1에 비해 컸다. 또한, 긍정적 정서 척도에서는 한국과 그 외 국가 간 θ_g의 차이에 비해, 한국과 그 외 국가 간 θ_1의 차이가 큰 반면에, 부정적 정서 척도에서는 한국과 그 외 국가 간 θ_g의 차이에 비해, 한국과 그 외 국가 간 θ_1의 차이가 작았다. 즉, 긍정적 정서 척도와 달리, 부정적 정서 척도에서는 ERS를 통제한 후에 한국과 그 외 국가 간 주관적 안녕감에 대한 잠재변인 추정치의 차이가 줄어들었다.

제 2 절 제한점 및 제언

본 연구의 제한점과 그에 대한 후속 연구에 대한 제언은 다음과 같다.

첫째, 본 연구에서는 ERS에 대해 분석하고 국가별 차이를 비교했으나, ERS가 나타나는 원인과 국가별 ERS의 차이가 존재하는 이유에 대한 분석은 수행하지 못하였다. 따라서, ERS의 원인을 파악하기 위해 실험연구와 질적면담을 활용한 추가 연구가 필요하다.

둘째, 본 연구에서는 IRTree 모형의 적합도를 파악하기 위해 다른 문항반응모형들과 적합도를 비교하고, 추정된 모수 간 상관을 분석하였으며, ERS 통제 전후의 주관적 안녕감에 대한 잠재변인 추정치 간 차이를 분석하였지만, 범주의 수, 범주의 제시 방법 등 여러 조건에 대한 IRTree 모형의 수행력을 살펴보지 못하였다. 따라서, 다양한 조건을 고려하여 IRTree 모형의 수행력을 파악할 수 있는 모의실험 연구가 필요할 것이다.

셋째, 본 연구에서는 ERS와 이를 통제한 주관적 안녕감에 대한 잠재변인만을 분석했으나, 다양한 응답양식을 동시에 추정하는 연구와 응답양식 뿐만 아닌 다른 응답 편파를 통제하기 위한 연구가 필요하다. 한 검사 내 두 가지 응답양식이 동시에 나타날 수 있다는 Weijter et al.(2010a)의 연구도 여러 응답양식을 동시에 통제하는 연구의 필요성을 뒷받침한다. 이를 위해 Plieninger & Heck(2018)이 제안한, MRS, ARS, ERS를 동시에 탐색할 수 있는 위계적 IRTree 모형을 활용할 수도 있을 것이다.

넷째, IRTree 모형을 사용하기 위해 다분문항반응을 가상의 이분문항반응으로 변환할 때, 다수의 결측치가 발생하지만, 이 결측치가 모수 추정에 미치는 영향에 대해서는 분석하지 못하였다. <표 3-2>에 제시되어

있듯이, 긍정적 ERS 혹은 부정적 ERS에 대한 가상의 이분문항에 모두 응답한 사례는 존재할 수 없으며, 두 개의 가상의 이분문항반응 중 하나는 결측치가 된다. 따라서, 가상의 이분문항반응을 생성할 때 발생한 결측치가 모수 추정에 미치는 영향에 대한 후속 연구가 필요하다.

참 고 문 헌

김석호, 신인철, 정재기. (2011). 응답자의 성격특성과 응답스타일. **조사연구, 12**(2), 51-76.

김정은, 이수란, 양수진. (2018). 한국형 그릿(Grit) 특성에 대한 고찰: 높은 그릿집단의 다차원적 유형 연구. **한국심리학회지: 문화 및 사회문제, 24**(2), 131-151.

박정. (1999). 다분 문항반응이론 모형의 능력모수 추정치의 편파도 감소를 위한 모수 추정방법. **교육평가연구, 12**(2), 195-218.

백순근, 채선희. (1998). **컴퓨터를 이용한 개별적응검사**. 서울: 원미사.

백순근, 박현정, 상경아, 임현수, 황은희, 정선혜, 김혜숙 (2007). **교육측정의 이론과 실제**. 파주: 교육과학사

성태제. (2019). **현대교육평가**. 서울: 학지사.

손원숙. (2015). 확인적 잠재계층 요인분석을 활용한 응답편파모형 탐색. **교육평가연구, 28**(1), 97-115.

신숭배. (2018). 한국 청소년 삶의 질 결정요인. **사회과학연구, 29**(1), 195-217.

장윤선, 김미림, 이주연. (2019). 문항반응나무(IRTree) 모형: 자기보고식 서열척도 자료 분석의 대안적 방법. **교육평가연구, 32**(2): 303-323.

정혜원, 박소영, 김정인, 김아름. (2021). 청소년의 읽기 소양과 삶의 만족도의 영향변인 탐색: PISA 2018 한국 핀란드 국제비교. **교육과정평가연구, 24**(1), 123-152.

조성민, 구남욱, 김현정, 이소연, 이인화(2019). **OECD 국제 학업성취도**

평가 연구: PISA 2018 결과 보고서. 한국교육과정평가원.

조영진, 이현숙. (2022). IRTree 모형을 활용한 TIMSS 2019 자료 내 ERS 탐색. **교육평가연구, 35**(1), 189-218.

Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences, 40*, 1235-1245.

Ames, A. J., & Leventhal, B. C. (2021). Modeling changes in response style with longitudinal IRTree models. *Multivariate Behavioral Research*, Advance online publication. DOI: 10.1080/00273171.2021.1920361

Batchelor, J. H., & Miao, C. (2016). Extreme response style: A meta-analysis. *Journal of Organizational Psychology, 16*(2), 51-62.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57,* 289-300.

Benitez, I., He, J., Van de Vijver, F. J., & Padilla, J. L. (2016). Linking extreme response style to response processes: A cross-cultural mixed methods approach. *International journal of psychology : Journal international de psychologie, 51*(6), 464-473.

Bock, D. (1972). Estimating item parameters and latent ability when responses are socred in two or more nominal categories,

Psychometricka, 37, 29-51.

Bolt, D. M., & Johnson, T. R. (2009). Addressing score bias and differential item functioning due to individual differences in response style. *Applied Psychological Measurement, 33*(5).

Böckenholt, U. (2017). Measuring response styles in Likert items. *Psychological Methods, 22*(1).

Böckenholt, U., & Meiser, T. (2017). Response style analysis with threshold and multi-process IRT models: A review and tutorial. *British Journal of Mathematical and Statistical Psychology, 70*(1), 159-181.

Böckenholt, U. (2019). Assessing item-feature effects with item response tree models. *British Journal of Mathematical and Statistical Psychology, 72*(3), 486-500.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment, *Journal of Statistical Software, 48*(6), 1 - 29.

Chen, C., Lee S.Y., & Stevenson. H. W. (1995). Response style and cross-cultural comparisons of rating scales among east asian and north american students. *Psychological Science, 6*(3), 170-175.

Clefberg Liberman, L., Larsson, K., Altuzarra, M. P., Ost, L., & Ollendick, T. (2015). Self-reported life satisfaction and response style differences among children in Chile and Sweden. *Journal of Child and Family Studies, 24*.

de Ayala, R. J. (2009). *The theory and practice of item response*

theory. New York: Guilford Press.

De Boeck, P., & Partchev, I. (2012). IRTrees: Tree-based item response models of the GLMM family. *Journal of Statistical Software, 48*, 1-28.

Diener, E., & Emmons, R. (1985). The independence of positive and negative affect. *Journal of Personality and Social Psychology, 47,* 1105-1117.

Eid, M. & Rauber, M. (2000). Detecting measurement invariance in organizational surveys. *European Journal of Psychological Assessment, 16*, 20-30.

Gilman, R., Huebner, E. S., Tian, L., Park, N., O'Byrne, J., Schiff, M., Sverko, D., & Langknecht, H. (2008). Cross-national adolescent multidimensional life satisfaction reports: Analyses of mean scores and response style differences. *Journal of Youth and Adolescence, 37*(2).

Greenleaf, E. (1992). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research, 29*, 176-188.

Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology, 20*(3), 296 - 309.

Jeon, M., & De Boeck, P. A. (2016). A generalized item response tree model for psychological assessments. *Behavior Research Method 48.*

Jeon, M., & De Boeck, P. (2019). Evaluation on types of invariance in

studying extreme response bias with an IRTree approach. *British Journal of Mathematical and Statistical Psychology, 72*, 517-537.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. New York: The Gilford press.

Lee, C., & R. T. Green. (1991). Cross-cultural examination of the Fishbein behavioral intentions model. *Journal of International Business Studies, 2,* 289-305.

Lee, S., Vasquez, E., Ryan, L., & Smith, J. (2020). Measurement equivalence of subjective well-being scales under the presence of acquiescent response style for the racially and ethnically Ddverse older population in the United States. *Survey Research Methods, 14*(4), 417-437.

Lee, S., & Lee, D. K. (2018). What is the proper way to apply the multiple comparison test?. *Korean journal of anesthesiology, 71*(5), 353 - 360.

Masters, G. N. (1982). A Rasch model for partial credit socring. *Psychometricka, 47*, 149-174.

McKnight, P.E. and Najab, J. (2010). Mann-Whitney U Test. In *The Corsini Encyclopedia of Psychology* (eds I.B. Weiner and W.E. Craighead). New Jersey: Wiley

Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education?. *Personality and Individual Differences, 44*(7), 1539-1550.

Muraki, E. (1992). A generalized partial gredit model: An application

of an EM algorithm. *Applied Psychological Measurement, 16*, 159-176.

OECD (2019). *PISA 2018 Results: What School Life Means for Student's Lives (Vol. Ⅲ)*. Paris: OECD Publishing.

Park, M., & Wu, A. D. (2019). Item response tree models to investigate acquiescence and extreme response styles in likert-type rating scales. *Educational and Psychological Measurement, 79*(5), 911-930.

Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and structure. *Personality and Social Psychology Bulletin, 21*, 100-108.

Penfield, R. D. (2014). An NCME instructional module on polytomous item response theory models. *Educational Measurement: Issues and Practice, 33*(1), 36-48.

Plieninger, H., & Heck, D. (2018). A new model for acquiescence at the interface of psychometrics and cognitive psychology. *Multivariate Behavioral Research, 53*(5).

Rost, J. (1991). A logistic mixture distribution model for polychotomous item resposes. *British Journal of Mathematical and Statistical Psychology, 44*(1), 75-92.

Samejima, F. (1969). Estimation of laten ability using a response pattern of graded scores, *Psychometrika Monograph Supplement, 34*, 100.

Soland, J., & Kuhfeld, M. (2020). Do response styles affect estimates

of growth on social-emotional constructs? Evidence from four years of longitudinal survey scores. *Multivariate Behavioral Research*, *56*(6), 853-873.

Verhoeven, K. J., Simonsen, K. L., & McIntyre, L. M. (2005). Implementing false discovery rate control: increasing your power. *Oikos, 108,* 643-647.

Warr, P. B., Barter, J., & Brownbridge, G. (1983). On the independence of positive and negative affect. *Journal of Personality and Social Psychology, 44*(3), 644 - 651.

Watson, D., & Clark, L. A. (1997). Measurement and mismeasurement of mood: Recurrent and emergent issues, *Journal of Personality Assessment, 68*(2), 267-296.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010a). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing, 27,* 236 - 247.

Weijters, B., Geuens, M., & Schillewaert, N. (2010b). The stability of individual response styles. *Psychological Methods, 15*(1).

Wetzel, E., Carstensen, C. H., & Bohnke, J. R. (2013). Consistency of extreme response style and non-extreme response style across traits. *Journal of Research in Personality, 47,* 178-189.

Zhang, Y., & Wang, Y. (2020). Validity of three IRT models for measuring and controlling extreme and midpoint response styles. *Frontiers in Psychology, 11,* 271.

Abstract

Analysis of Extreme Response Style to PISA 2018 Subjective Well-being Scale Using IRTree Model

Hyun-Jee Oh

Department of Education

The Graduate School

Seoul National University

The purpose of this study is to examine whether the Item Response Tree model (IRTree model) is appropriate to estimate the extreme response style (ERS) to the PISA(Programme for International Student Assessment) 2018 subjective well-being scale, the level of latent variables for subjective well-being after controlling the ERS, and to compare the difference between the level of ERS of Korea and that of the OECD (Organization for Economic Cooperation

and Development) countries except Korea. This study also aims to analyze the change in the level of latent variables for subjective well-being in Korea after controlling the ERS. According to PISA 2018 results, Korean students reported relatively higher subjective well-being compared to life satisfaction. Previous studies have observed self-report tests such as PISA 2018 are vulnerable to response bias such as ERS. Previous studies on the wellbeing of adolescents have observed differences in varied response styles between respondents when answering queries on wellbeing, and ERS distinction were the most reported. Additionally, scholars have found a relative increase in ERS levels when the number of response categories is small, such as on a 4 point scale like subjective well-being scale on PISA 2018. Such distortions result in differences in results reported by different cultural groups notwithstanding the traits of interest. Therefore, this study aims to determine whether Korean students reported relatively higher subjective well-being compared to life satisfaction because of high ERS levels in responding to the subjective well-being scale. This study analyzed data of 64,000 students randomly sampled by 2,000 students from 32 OECD countries who responded to the PISA 2018 subjectvie well-being scale.

The research questions are as follows.

Research question 1. Is the IRTree model appropriate to estimate the ERS to PISA 2018 subjective well-being

scale, and the level of latent variables for subjective well-being after controlling the ERS?

Research question 2. How are Korea's ERS estimates different from those of other OECD countries?

Research question 3. How does the level of latent variable for subjective well-being change in Korea after controlling the ERS?

The major outcomes of this study are noted as follows.

First, ERS was observed in the PISA 2018 subjective well-being responses of Korean students. The IRTree model was most suited to this analysis. In addition, the restricted IRTree model, which assumes no direction of ERS(positive or negative), was found inapt for the data. Thus, directionality was observed in the ERS.

Second, the reliability of the IRTree model was demonstrated through correlation analyses between parameters estimated through the Graded Response Model(GRM) and the IRTree model. A high positive correlation was found between the estimates of latent variables for subjective well-being before and after controlling ERS. Additionally, the correlation coefficient between estimates of latent variables for subjective well-being of the two scales(postive emotion and negative emotion scales) remained constant before and after controlling the ERS. A high positive correlation was found with estimates of latent variables for subjective well-being after controlling

the ERS in instances of estimates of positive ERS to the positive emotion scale, estimates of the negative ERS, and estimates of positive ERS to the negative emotion scale.

Third, it was confirmed that Korean students tended to respond more extremely than their counterparts in other OECD countries. The results of the Mann-Whitney U test, the Kruskal-Wallis H test, and the post-hoc test evinced that Korean students displyed greater positive ERS esitmates for the positive emotion scale, and both positive and negative ERS estimates for the negative emotion scale than students in other countries.

Fourth, the ERS significantly affected the estimates of latent variables for subjective well-being of Korean students. The Wilcoxon Singned Rank Test demostrated that the estimates of latent variables for subjective well-being increased significantly on the positive motion scale after controlling the ERS. However, the estimates of latent variables for subjective well-being decreased significantly on the negative emotion scale after controlling the ERS.

Fifth, it can be inferred that the ERS contributed to estimates of latent variables for the subjective well-being of Korean students being greater than those of students in other OECD countries on the negative emotion scale. The differences in the estimates of latent variables for subjective well-being decreased between Korea and other countries after controlling the ERS from the negative emotion scale.

This study is significant in that it empirically confirmed the

influence of the ERS using large-scale international data. Further, it demonstrated the utility of the IRTree model for the ananlysis of ERS. In future studies, experimental studies and qualitative interviews will be needed to determine the cause of the difference in ERS. In addition, a simulation study is needed to analyze the performance of the IRTree model. Prospective research projects should also simulaneously estimate varied response styles and analyze the effects of missing values in pseudo dichotomous responses on paraemter estimation.

keywords : PISA 2018, subjective well-being, extreme response style, IRTree model

Student Number : 2020-26211