

## 수학 학업성취도가 높은 국가의 수학-정의적 영역 요인 분석 및 측정 동일성 검증<sup>1)</sup>

이 종 희\* · 김 기 연\*\* · 김 수 진\*\*\*

수학 학습과 성취에 있어서 인지적 영역뿐만 아니라 정의적 영역의 여러 요인이 영향을 주고 있다는 연구가 활발하게 진행되고 있으며 교육 활동에 있어서 정의적 영역에 대한 중요성 인식은 국가수준의 교육과정 개정이나 학업성취도 평가, 국제비교평가 연구 등에서도 확인해 볼 수 있다. 이와 같이 수학 학습과 성취에 영향을 주는 요인으로서 정의적 요소를 살펴보기 위한 다양한 연구 방법 중에 대규모의 표집 집단을 대상으로 하는 경우 가장 보편적이고 일반적으로 사용되는 것이 설문에 따른 반응을 분석하는 것이다. 이에 본 연구에서는 PISA 2003과 TIMSS 2007의 자료를 분석하여 수학 학업 성취도가 상대적으로 높은(상위 30%) 국가를 선정하고 각 국가의 정의적 영역 성취점수를 분석함으로써 국제비교평가의 결과를 바탕으로 한 우리나라 학생들의 정의적 특성을 잘 파악할 수 있는지를 알아보았다. 이 결과를 바탕으로 우리나라 학생들의 정의적 특성을 파악하기 위한 설문 문항 개발의 필요성을 확인하였고 대규모 국제비교평가의 설문문항이 우리나라 학생들을 대상으로 하는 설문 문항 개발에 어떤 시사점을 줄 수 있는지 고찰하였다.

### 1. 서론

교육과정에 제시된 수학과목의 성격에서는 수학적 지식과 기능을 활용하여 실생활의 여러 가지 문제를 해결해 봄으로써 수학의 필요성과 유용성을 인식하고, 수학 학습의 즐거움을 경험함으로써 수학에 대한 긍정적인 태도를 가지도록 할 것을 강조하고 있으며, 수학과목의 하위 목표 중 하나는 수학의 가치를 이해하여 수학에 대한 관심과 흥미를 지속적으로 가지며, 수학에 대한 긍정적 태도를 기른다는 것이다(교육과학기술부, 2007). 이처럼 국가수준의 교육과정에서 가치 인식이나 긍정적인 태도, 관심과 흥미 등의 정의적인 영역에 대한 요소를 제

시하고 있다는 것은 학생들이 수학을 학습함에 있어서 인지적인 영역의 능력 못지않게 정의적 영역의 요소에 의해서도 많은 영향을 받는다는 것을 주목하고 있다는 것이며 수학에 대한 긍정적 정의(affect)를 형성하는 것이 수학교육의 목표 중 하나임을 강조하는 것이라 할 수 있다. 국가수준의 학업성취도 평가나 TIMSS, PISA 등과 같은 국제비교평가와 같이 학생들의 수학 학업성취도에 대한 연구에서도 학생들의 성취도에 영향을 주는 여러 가지 배경 요인을 조사하는 데에 있어 정의적 영역의 요소에 대한 것이 포함된다는 것도 같은 맥락에서 생각해 볼 수 있을 것이다.

국제비교평가를 통해 본 우리나라 학생들의 수학 학업성취도는 매우 높은 수준이라고 할

\* 이화여자대학교 (jonghee@ewha.ac.kr)

\*\* 이화여자대학교 (freenego@lycos.co.kr)

\*\*\* 한국교육과정평가원 (sjkim@kice.re.kr), 교신저자

1) 이 논문은 한국연구재단의 지원(KRF-2009-32A-B00216)에 의하여 이루어진 연구임

수 있다. TIMSS 2007에서 우리나라 중학교 2학년 학생들의 수학 학업성취도는 50개 참여국 중에 2위를 차지하였고, TIMSS 2003에 비해 통계적으로 유의한 차이로 8점 상승하였다. TIMSS 1995에서 국제 순위 3위를 기록한 이후로 TIMSS에서 우리나라의 수학 학업성취도 순위는 2위를 유지하고 있으며 성취도 점수 또한 증가하는 추세다(박정, 정은영, 김경희, 한경혜, 2004; 김경희, 김수진, 김남희, 박선용, 김지영, 박효희, 정송, 2008a). 만 15세 학생들을 대상으로 수학적 소양을 평가하는 PISA 2006의 결과에서도 우리나라는 OECD 국가를 대상으로 하였을 때 1~2위, 전체 참여국 중에서는 1~4위를 차지하여 높은 수학적 소양을 유지하고 있으며, PISA 2003에 비해서도 5점이 향상된 점수였다. 하지만 이들 연구 결과에는 우리나라 학생들의 정의적 영역에 대한 성취<sup>2)</sup>가 국제적으로 매우 낮은 수준에 있다는 것을 보고하고 있다.

높은 학업성취에 비해 정의적 영역의 성취가 상대적으로 낮음에도 불구하고 이와 같은 현상에 대한 분석이나 해석을 위한 연구는 아직 미진한 상황이다. 이에 본 연구에서는 수학 학습과 성취에 관련된 정의적 요인을 살펴보기 위한 설문 연구법에 대하여, 우리나라 학생들이 보여주는 특징이나 현상을 알아보기 위한 도구로서 어떤 요건을 갖추어야 할 것인지를 논의할 것이다. 이를 바탕으로 효과적으로 측정해 낼 수 있는 설문 도구를 개발하기 위한 선행연구로서 우선 측정 도구의 신뢰도와 타당도, 문항 합치도 등을 검증받은 대규모 국제비교평가(PISA, TIMSS)의 정의적 영역에 대한 설문 문항을 세 가지 방법으로 분석하고자 한다. 대규모의 국제비교평가

인 PISA 2003과 TIMSS 2007에서 높은 수준의 성취를 보인 국가들을 대상으로 PISA와 TIMSS의 본래 의도대로 정의적 요인들<sup>3)</sup>이 잘 나타나는 개념들인지 각 국가별로 탐색적 요인분석을 실시한다. 또한 확인적 요인분석을 실시한 후, 7개의 국가<sup>4)</sup>를 통합하여 분석하는 다집단 분석을 실시하여 각 국가들의 요인구조가 같다고 할 수 있는지 확인해 보고자 한다. 이 결과를 바탕으로 우리나라 학생들의 수학 성취도에 영향을 미치는 정의적 영역 요소를 조사하기 위한 설문 문항 개발의 필요성을 재고(再考)하고 우리나라 학생들에게 적합한 문항 개발에 대한 시사점을 논의할 것이다.

## II. 이론적 배경

### 1. 설문지를 통한 정의적 영역의 측정

수학 학습과 성취에 관련하여 정의적 요인을 살펴보기 위한 효율적인 방법 중 하나로서 설문지를 개발하여 학생들의 정의적인 면을 파악하는 것이 있다. 학생들의 수학에 대한 정의적 영역을 탐색하기 위한 연구로는 개인 수준의 연구를 비롯하여 국가수준, 국제 비교 연구 수준 등 다양하게 이루어지고 있다. 국내에서도 수학 학습이나 평가에 있어서 정의적 요인의 영향을 측정하거나 판단하려는 시도가 이루어져 왔으며 PISA, TIMSS와 같이 여러 국가의 학생들을 대상으로 한 국제 비교 연구에서도 설문을 통한 정의적 영역의 측정이 이루어져 오고 있다.

2) 이 연구에서는 정의적 영역의 성취를 동기, 흥미, 자아개념 같은 학생들의 정의적 영역에서의 점수를 의미하는 것으로 사용한다.

3) PISA는 수학에 대한 흥미, 도구적 동기, 자아효능감, 불안, 자아개념의 5개 요인, TIMSS는 수학 학습의 즐거움 인식, 수학 학습에 대한 자신감, 수학학습에 대한 가치 인식의 3개 요인을 측정한다(OECD,2003).

4) PISA 2003에서는 호주, 캐나다, 핀란드, 홍콩, 일본, 한국, 뉴질랜드 TIMSS 2007에서는 영국, 홍콩, 일본, 한국, 싱가포르, 대만, 미국의 7개 국가의 자료를 분석하였다.

김부미(1996)는 기존에 개발되어 있는 여러 가지 설문지를 바탕으로 수학적 태도에 대한 교사와 학생 변인을 알아보는 검사 도구를 제작하였으며, 이경희(2003)는 정의적 요인이 학생들의 학업성취 수준이나 학년에 따라 어떤 특성을 나타내는지를 살펴보기 위해 귀인, 자아개념, 수학 불안 요인 등을 측정하는 검사지를 활용하였고 신성균 외(1992)는 수학과 평가도구를 개발함에 있어 정의적 영역 평가 도구의 필요성을 강조하고 수학적 성향 검사 도구를 개발하였다. 수학생태와 같이 특정한 범주의 학생들의 수학에 대한 정의적 특성을 알아보기 위한 연구에서도 설문을 통한 검사가 활용되었다. 김선희 외(2005)는 수학 영재와 과학영재 그리고 일반 학생들의 인지적·정의적·정서적 특성을 비교하는 연구에서 수학적 성향과 자기 효능감을 알아보는 설문 조사를 실시하였으며 강순자 외(2006)는 수학 영재의 수학에 대한 정의적 특성을 살펴보기 위해 정의적 특성의 하위 요소를 학문적 자아개념, 태도, 흥미, 수학 불안, 학습 습관의 다섯 가지로 구분하고 이를 측정하기 위해 검사 도구를 제작하였다. 자아개념은 자기 자신에 대한 스스로의 평가로 자기 자신이 어떤 능력을 가지고 있으며 그 수준이 얼마나 되는가를 스스로 평가하는 것과 자기 자신에 대한 애정과 존중감을 포함한다. 태도, 특히 교과(학습)에 대한 태도는 그 교과의 성취나 호, 불호의 판단과도 연관되는 것으로서 이영주(1999)에 의하면 학교 학습에 대한 질적 태도는 학습자의 질적 자아개념과 유의적인 상관관계가 있으며 학업성적이 낮은 학생은 자아개념이 낮은 경향을 보이면서 학교에 대한 질적 태도를 나타낸다(강순자 외, 2006, 재인용). 그 외에도 지윤정(2004)은 설문 조사와 교수 실험을 통해 학생들의 국제 비교 평가에서 우리나라 학생들의 수학 성취가 높음에도 불구하고 수학에 대한

흥미와 자신감이 낮은 것에 대한 원인을 탐색하였으며 윤나은(2005)은 수학 불안을 불안의 특수한 형태로, 일상생활이나 학교에서 수학문제를 해결하도록 요구받았을 때 일어날 수 있는 무기력, 공포, 마비, 정신적 혼란과 같은 신체·정서적 반응이라 정의하였으며 내용 불안과 평가 불안으로 구분하기도 하였다.

개인 수준의 연구는 설문 조사 대상의 범위가 제한적인 것에 비하여 국가 수준의 연구나 국제 비교 평가 연구에서는 같은 연구방법을 사용한다 하더라도 전수조사를 실시할 수 있고, 표집을 하는 경우에도 개인 연구의 조사 대상에 비해 대규모의 표본을 선정할 수 있으므로 연구 결과의 일반화가 더 용이하다. 즉, 국가 수준의 학업성취도 평가나 PISA, TIMSS 등의 대규모 연구는 개인 수준의 연구에 비해 국가 전체의 특성이나 성향을 거시적인 관점에서 보여줄 수 있다.

PISA 2003에서 조사한 정의적 영역은 수학에 대한 흥미, 수학에 대한 도구적 동기, 수학에 대한 자아 개념, 수학에 대한 자아효능감, 수학에 대한 불안감을 알아보는 설문 문항으로 구성되어 있는데(이미경 외, 2004), 각각의 문항은 다음과 같이 크게 네 가지의 영역으로 구분할 수 있다. 첫째, 동기유발과 관련된 요인 및 학교에 대한 일반적인 태도를 묻는 영역으로 수학에 대한 흥미와 즐거움, 수학에 대한 도구적 동기, 학교에 대한 태도, 학교에 대한 소속감을 묻는 문항이 포함되어 있다. 둘째, 수학에 대한 학습자 관련 신념을 묻는 영역으로 수학에 대한 자기 효능감, 자아개념을 묻는 문항이 포함된다. 셋째, 수학에 대한 감정적 요인을 묻는 영역으로 여기에는 수학에 대한 불안을 묻는 문항이 포함된다. 끝으로 수학에 대한 학습 전략을 묻는 영역으로 여기에는 기억이나 연습 전략, 정교화 전략, 통제 전략 등에 대한 문항이 포함되어 있다. 그리

고 TIMSS 2007은 수학에 대한 자신감, 수학 학습에 대한 즐거움 인식, 수학에 대한 가치 인식을 알아보는 문항으로 구성된(김경희 외, 2008a) 설문지를 통해 학생들의 수학에 대한 정의적 성취를 측정하고자 하였다. 끝으로 우리나라의 국가수준 학업성취도 평가는 교과목의 학습에 대한 문항을 통해 학생들의 수학적 흥미, 불안, 자신감 등을 파악하고자 하였다(양길석 외, 2007; 김성숙 외, 2009).

이와 같이 검사 도구를 통해 수학 학습에서의 정의적 특성을 파악하려는 연구에서는 그 하위 요소를 무엇으로 규정할 것이며 어떤 검사 도구를 제작 또는 선정할 것인가가 매우 중요하다고 할 수 있다. 특히 우리나라에서도 국가 수준의 연구를 통해 우리나라 전체 학생들의 특성과 수준을 파악하는 평가가 활발하게 시행되고 있기 때문에 연구의 목적과 규모에 알맞은 검사 도구의 필요성은 강조되어야 한다. 현재 우리나라의 국가수준 학업성취도 평가에서도 각 교과목의 학습에 대한 문항을 통해 학생들의 수학적 흥미, 불안, 자신감 등을 파악하고 있으나(김성숙 외, 2009) 일부 표본을 대상으로 하여 실시하던 것을 2007년 이후부터 전수 평가로 전환하면서 보다 정확한 결과 도출과 분석이 가능해 졌다고 볼 수 있다. 그러나 오히려 정의적 영역을 측정하기 위한 문항의 수가 줄었기 때문에 커진 표본의 크기만큼의 효과를 기대하기가 어려워 질 수도 있다. 따라서 전국의 모든 학생들을 대상으로 한 평가에서 활용하기 위해서는 정의적 영역에 대한 문항을 좀 더 보완하고 수정해야 할 필요가 있다. 그러나 새로운 설문 도구 개발에 앞서 기존에 활용된 문항을 면밀하게 분석함으로써 우리나라 학생들의 높은 학업 성취와 상반되는 정의적 영역의 특징에 대한 원인 규명과 진단을 내릴 수 있어야 할 것이다.

## 2. 구인 확인 및 측정동일성 검증

설문 도구가 구성이 되면 설문 도구 내에 포함된 설명되지 않은 인간의 심리적 특성이나 성질을 심리적 구인으로 분석하여 조작적 정의를 부여하여야 한다. 이렇게 조작적으로 정의된 심리적 구인들을 검사 점수를 통해 표현하는데 검사 점수가 이러한 심리적 구인들을 제대로 측정하고 있는지를 검증하는 방법 중 하나가 구인타당도(construct validity)이다(성태제, 2002). 구인 타당도를 검증하는 통계적 방법 중 가장 대표적인 것이 요인분석(factor analysis)이라고 할 수 있다. 요인분석이란 여러 변수 사이에 존재하는 상호관련성을 찾아내어 소수의 요인으로 추출하여 요인들의 의미를 부여하는 통계적인 방법이다. 성태제(2002)는 요인분석을 통한 구인타당도를 검증하는 기본 절차를 간단히 설명하였는데, 우선 필요한 문항을 제작하여 검사를 실시한 후에 문항점수를 얻는 것이 기본이 된다. 다음으로는 문항들 간의 상관계수 행렬을 구한 후 회전하지 않은 요인을 추출한다. 그리고 요인들을 회전하여 다시 추출해 본다. 회전된 요인에서 관계가 있는 요인에 해당하는 문항들의 내용에 근거해서 요인의 이름을 붙이고 이를 해석한다.

요인분석을 통해 설문 도구의 구인타당도를 검증할 수 있다. 구인타당도는 피험자들의 응답 자료를 가지고 통계적 방법으로 검증되므로 과학적이고 객관적이다(성태제, 2002). 요인분석을 통해 구인타당도를 검증하게 되면, 여러 심리적 특성들에 대한 조작적 정의를 설정하게 되고 앞에서 설명한 통계적 방법 및 절차에 따라 심리적 특성을 측정하는 문항들의 관련성을 통해 이러한 정의의 타당성을 밝혀주기 때문에 설문 도구에 대한 타당성에 근거가 된다.

다집단 분석은 서로 다른 모집단 즉, 다른 특성을 가진 집단의 자료들을 모형에 합치시켜 봄

으로써, 집단 간 비교를 하고자 하는 목적으로 실시된다(Byrne, Shavelson, & Muthen, 1989; Joreskog, 1971). 집단 간 비교 검증을 하기 위해서는 우선 집단 간 측정 동일성이 충족되는지 확인하는 작업이 필수적이다. 측정 동일성을 검증하는 방법론으로 가장 많이 쓰이는 방법론은 구조방정식이다(예: Byrne, Shavelson, & Muthen, 1989; Joreskog, 1971; Meredith, 1993; Reise, Widaman, & Pugh, 1993; Steenkamp & Baumgartner, 1998; Vandenberg, 2002; Vandenberg & Lance, 2000). 구조방정식을 이용한 측정 동일성 검증(Measurement Equivalence/Invariance, ME/I)을 하면, 관심 있는 부분의 모수 제약을 통해 집단 간 어떤 부분 예를 들면 측정틀, 측정단위, 측정원점, 요인분산 혹은 공분산, 측정오차분산 등에서 차이가 있는지 알아낼 수 있다. 측정 동일성 검증을 통해 집단 간 측정 동일성이 성립되면, 집단 간에 요인구조가 같아 측정하는 것이 같다고 할 수 있다. 그러므로 집단 간 측정 동일성이 성립해야만 진정한 의미에서 집단 간 비교가 가능하다고 할 수 있다. 반대로 측정 동일성이 만족되지 않는다면, 집단 간의 차이가 실제 집단 간 속성 차이인지, 검사 도구에 의해 나타난 차이인지 알 수 없게 된다(Millsap & Kwok, 2004).

연구의 목적에 따라 다르지만, 보통 측정 동일성 검증을 하는 순서는 다음과 같다(Vandenberg & Lance, 2000). 먼저 측정틀 동일성(configurational invariance)을 검토한 후, 측정단위 동일성(metric invariance), 측정원점 동일성(scalar invariance), 요인분산/공분산 동일성(invariant factor variance/covariance), 측정오차분산 동일성(invariant uniqueness) 순서로 검증을 실시한다. 보통 측정원점 동일성 검증까지는 이루어져야 두 집단 간 비교가 가능하다고 할 수 있다. 나머지 단계의 검증들은 연구의 목적에 따라 이루어지기도 하고 생략되기도 한다

(Vandenberg & Lance, 2000). 각 단계에서의 검증에 대한 간단한 설명을 아래에 제시한다.

먼저, 측정틀 동일성(configurational invariance)은 각 집단마다 문항들을 인지하는 틀이 같은 것을 의미한다. 즉, 요인구조가 같다는 뜻이다. 그러나 이때 전체적인 틀만 같은 것이지, 요인계수나 기타 모수들은 집단 간에 다를 수 있다. 두 번째로, 측정단위 동일성(metric invariance)은 각 요인에 해당하는 문항들이 기여하는 정도가 집단 간에 같은 것을 의미한다. 구조방정식에서 요인계수들이 같다고 동일화 제약을 함으로써 측정단위 동일성 검증을 해볼 수 있다. 세 번째, 요인분산 동일성(invariant factor variance)은 각 요인의 분산이 동일한지 검증해보는 것이다. 요인분산 동일성과 측정오차분산 동일성이 동시에 만족되면, 고전검사이론의 신뢰도<sup>5)</sup>의 개념에 의거하여 집단 간 신뢰도가 같다고 결론내릴 수 있다. 또한 요인분산과 요인공분산 동일성이 동시에 만족되면 요인 간 상관관계가 집단 간에 같다는 결론을 내릴 수 있다. 네 번째, 요인공분산 동일성(invariant factor covariance)은 각 집단의 요인 간 공분산이 동일한 것을 의미한다. 위에서 말한 것처럼, 요인분산 동일성과 동시에 만족되면 모든 집단에서 요인 간 상관계수가 같다고 할 수 있다. 다섯 번째, 측정오차분산 동일성(invariant uniqueness)은 집단 간 측정오차분산들이 동일한 것을 나타낸다. 아직까지는 측정오차분산 동일성을 독립적으로 해석하는 경우는 없으며, 요인분산 동일성과 동시에 만족된다면, 집단 간 신뢰도가 같다는 결론을 내릴 수 있다. 마지막으로, 측정원점 동일성(scalar invariance)은 집단 간 절편이 같은 것을 의미한다. 구조방정식에서는 공분산구조와 평균구조, 두 가지 구조가 분석될 수 있다. 공분산 구조에서는 절편의 개념이 없지만, 평균구

5) (신뢰도) =  $\frac{(\text{진접수의 분산})}{(\text{측정변수의 분산})}$

조(mean structure)에서는 측정 원점이라고 할 수 있는 절편의 개념이 존재한다. 측정원점 동일성에서는 절편이 집단 간에 같다고 동일화 제약을 해주게 된다. 측정원점 동일성이 만족 되면, 집단 간에 반응 식역이 같은 것으로 해석할 수 있다. 즉, 리커트 척도를 예로 들자면, 모든 집단 간에 1-5까지의 느껴지는 정도가 같다고 할 수 있다. 반대로 측정원점 동일성이 만족하지 않는다면, 어떤 집단에서는 2점을 매우 낮은 수준으로 인식하는 한편, 다른 집단에서는 2점을 그보다는 높은 수준으로 인식할 수도 있다. 그러므로 측정원점 동일성이 충족되어야 진정한 의미에서 집단 간 비교가 가능하다고 할 수 있다.

### III. 연구 방법

#### 1. 분석 자료 및 분석 대상

수학 학습과 성취에 영향을 미치는 정의적 요인을 알아보기 위한 설문 연구 도구 및 자료를 분석하기 위해 공개되어 있는 자료 중 수학에 대한 문항이 가장 많이 포함되어 있는 PISA 2003과 TIMSS 2007의 자료를 선정하였다.

##### 가. PISA 2003

PISA의 경우, PISA 2003 참여국 중 수학 성취도가 높은 국가를 선정하였다. PISA 분석에 포함된 국가는 우리나라를 비롯하여 호주, 캐나다, 핀란드, 홍콩, 일본, 뉴질랜드의 7개국이다. PISA에 포함된 정의적 영역으로는 수학에 대한 흥미와 즐거움, 수학에서의 도구적 동기, 수학에 대한 자아효능감, 수학에 대한 불안감, 수학에 대한 자아개념이 있다. PISA 2003의 정의적 영역에 해당하는 요인은 수학에 대한 흥미와 즐거움, 수학에서의 도구적 동기, 수학에

대한 자아효능감, 수학에 대한 불안감, 수학에 대한 자아개념의 5개(OECD, 2004, 이미경 외, 2004)인데 본 연구에서는 이를 각각 요인P-1, P-2, P-3, P-4, P-5로 표기할 것이다.

##### 나. TIMSS 2007

TIMSS의 경우 TIMSS 2007 참여국 중 수학 성취도가 높으면서 정의적 영역에서도 높은 점수를 보이는 국가를 선정하였다. TIMSS 분석에 포함된 국가는 우리나라를 비롯하여 영국, 홍콩, 일본, 싱가포르, 타이완, 미국의 7개국이다. TIMSS에 포함된 정의적 영역으로는 자신감지수, 즐거움 인식지수, 가치인식 지수 등이 있으며 이 연구에서는 3가지 항목에 TIMSS 2007의 정의적 영역에 해당하는 요인은 수학학습의 즐거움 인식, 수학학습에 대한 자신감, 수학 학습에 대한 가치 인식인데(Mullis, Martin, Ruddock, O'Sullivan, Arora, & Erberber, 2005; 김경희 외, 2008b) 본 연구에서는 이를 각각 요인 T-1, T-2, T-3으로 표기할 것이다.

#### 2. 분석 방법 및 절차

각 국가별로 구인타당도를 검증하기 위하여 우선 탐색적 요인분석을 실시하였고, PISA와 TIMSS에서 나타내는 이론에 근거한 확인적 요인분석을 실시하였다.

##### 가. 구인타당도 검증을 위한 탐색적 요인분석 및 확인적 요인분석

PISA와 TIMSS의 문항들의 구성 요소들을 확인하기 위하여 각 국가들을 대상으로 PISA나 TIMSS에서 원래 의도했던 대로 정의적 요인들이 잘 나타나는 개념들인지 탐색적 요인분석을 실시하였다. 탐색적 요인분석을 위해 SPSS 17.0 소프트웨어를 이용하였다. 요인추출 방법은 주축 요인 방법(principle factor extraction)을 사용했고, 해석을 위해 회전은 요인 간 상관의 최

대로 있도록 하는 직접 오블리민(direct oblimin,  $\delta=0$ )을 사용했다.

또한, 각 국가들을 대상으로 PISA에서 원래 의도했던 대로 정의적 요인들(수학에 대한 흥미, 도구적 동기, 자아효능감, 불안, 자아개념)이 잘 나타나는 개념들인지 확인적 요인분석을 하였다. 확인적 요인분석을 위해 AMOS 7.0 소프트웨어를 사용했고, 각각 PISA와 TIMSS에서 원래 의도했던 5개의 요인과 3개의 요인이 잘 나타나는지 분석하였다.

#### 나. 측정 동일성 검증을 위한 다집단 분석

PISA 2003과 TIMSS 2007에서 국가들 간 측정 동일성을 확인하기 위해 AMOS 7.0으로 다집단 분석을 실시하였다. 측정 동일성의 순서는 측정틀 동일성, 측정단위 동일성, 요인분산 동일성, 요인공분산 동일성, 측정오차분산 동일성, 그리고 측정원점 동일성 검증의 순서로 실시했다.

측정틀 동일성 검증에서는 각 국가들의 자료를 확인적 요인분석 모형에 합치시키게 되는데 이 때 모수, 예를 들면 요인계수, 요인분산 또는 공분산, 측정오차분산, 절편 등에는 아무런 제약을 주지 않아, 각 국가별로 개별의 모수를 추정하게 된다.

측정단위 동일성 검증에서는 측정틀 동일성 검증 모형에서 국가 간 요인계수들이 같도록 제약한다. 그리고 이전 측정틀 동일성 검증 모형과의 합치도 차이를 통해 측정단위 동일성이 만족될 수 있는지 판단하게 된다. 이후 요인분산, 요인공분산, 측정오차분산 동일성 검증에서는 순서대로 요인분산, 요인공분산, 그리고 측정오차분산을 국가 간에 동일하도록 제약하면서 전 모형과의 비교를 통해 동일성을 검증하게 된다.

마지막으로 측정원점 동일성 검증에서는 평균구조를 분석하게 된다. 평균구조 분석 시, 절편과 요인평균을 추정하기 위해서는 임의로 모

수 중 하나를 집단 간에 동일하도록 고정하고, 나머지 모수는 하나의 집단을 기준으로 추정하게 된다. 여기서는 측정원점(절편)이 국가 간에 같도록 제약을 한 후, 우리나라의 요인 평균을 0으로 고정하고 나머지 국가들의 요인 평균을 상대적으로 추정한다.

## IV. 연구 결과

### 1. 탐색적 요인분석 결과

측정 구성 요소를 탐색적으로 확인해보기 위하여 탐색적 요인분석을 실시하였다. 요인추출 방법은 주축 분해 방법을 사용했고, 회전은 요인 간 상관관계가 있도록 하는 오블리민 회전을 하였다. PISA 2003의 경우는 요인 개수를 5개로 설정하였고, TIMSS 2007의 경우는 요인의 개수를 3개로 설정하여 결과를 확인하였다. 여기서 중요하게 참조하는 것은 요인계수의 크기가 적절할지이며, 요인계수가 너무 작으면 수렴 타당도가 낮다고 할 수 있다. 그리고 요인 간 상관관계가 너무 크게 나타나 요인 간 변별이 되지 않는지 확인하고자 한다. 요인을 구성할 때는 요인계수의 절댓값이 0.4이상인 경우를 기준으로 하여 요인을 추출하였으며 요인계수가 두 가지 요인에서 높게 나타날 경우는 두 요인 모두에 속한 것으로 보았다.

#### 가. PISA 2003

5개의 요인들이 대체적으로 잘 나뉜 국가는 호주였다. 요인계수는 .460~.925이었고 요인 간 상관관계는 -.483~.554로 적절하였다. 이에 비해 다른 국가들은 각기 요인이 달리 나뉘는 특성을 보였다. 캐나다는 수학에 대한 불안과 자아개념에 관한 문항들이 하나의 요인으로 추출되었

다. 자아효능감과 관련된 문항들이 두 개의 요인으로 갈라졌다. 즉, 자아효능감이 두 개의 요인으로 나뉘게 된 것이다. 요인계수는 -.828~.924이었고 요인 간 상관은 .308~.598로 적절하였다. 핀란드의 경우는 요인들이 대체적으로 잘 나뉘었으나 수학에 대한 불안과 자아개념에 대한 문항들이 서로 걸친 구조를 형성하고 있다. 즉, 어떤 문항들은 수학 불안과 자아개념 모두에 대해서 높은 요인계수값을 보인다. 요인계수는 -.888~.859이었고 요인 간 상관은 -.611~.652로 적절하였다. 홍콩은 대체적으로 요인들이 잘 나뉘었으나 수학에 대한 불안과 자아개념이 약간 겹치는 것으로 나타났다. 요인계수는 -.728~.838이었고 요인 간 상관은 -.598~.370으로 적절한 정도로 나타났다. 일본의 경우, 수학에 대한 불안과 자아개념 문항이 하나의 요인으로 묶여 나왔다. 그리고 자아효능감의 일부 문항들이 따로 묶였다. 요인계수는 -.798~.910이었고 요인 간 상관은 -.491~.646으로 적절하였다. 뉴질랜드는 대체적으로 잘 묶였는데, 자아개념과 관련된 문항 두 개가 수학불안과 함께 묶였으며 요인계수는 -.920~.858이었고 요인 간 상관은 -.565~.348로 적절하였다.

우리나라의 경우, 요인계수는 .301~.828이었고 요인 간 상관은 -.489~.553으로 적절하였다. 요인의 개수를 5개로 놓고 탐색해 본 결과, 첫 번째 요인인 요인 A에 해당하는 문항들은 수학에 대한 흥미(요인 P-1)와 자아개념(요인 P-5)에 해당하는 내용으로, PISA에서 이론적으로 두 개로 나뉘어져 있던 요인들이 합쳐져서 나타났다. 또한 PISA 2003에서 자기효능감(요인 P-3)과 관련된다고 한 문항들은 요인 B와 요인 D로 나뉘어져 두 개의 요인이 되었다. 나머지 도구적 동기(요인 P-2)와 수학에 대한 불안(요인 P-4)에 관한 문항들은 요인 C와 요인 D로 비교적 잘 나뉘어졌다. 우리나라의 경우에는 하나 이상의 요인과 비교적 높은 상관을 보이는 문항도 있었다. 수학에 대한 자아효능감을 묻

기 위한 문항 중 ' $3x+5=17$ 에서  $x$ 를 구하는 것'과 같은 일차방정식 문항, ' $2(x+3)=(x+3)(x-3)$ 에서  $x$ 를 구하는 것'과 같은 이차방정식 문항은 요인 A(흥미+자아개념)와도 높은 상관이 나타났는데, 이는 다른 설문 문항에 비해 직접적으로 수식을 제시하는 문제로서 답을 구해낼 수 있는 문제이기 때문에 문제를 해결할 수 있는지의 여부가 응답을 하는 데에 영향을 미칠 수 있다. 요인이 합쳐진다는 것은 해당 문항이 하나의 요인에만 속하는 것이 아니라 다른 요인에도 공통적으로 나타남을 의미한다. 요인이 나뉜다는 것은 예를 들면, 원래 연구자들의 이론에서는 수학에 대한 흥미요인이어야 하는데, 흥미 요인이 또 다른 두개의 다른 개념으로 나뉘는 것을 의미한다.

탐색적 요인분석 결과를 종합해 보면, 원래 의도했던 대로 정의적 영역의 요인들이 가장 잘 나뉜 국가는 호주였다. 즉, 하나의 문항이 묻고자 하는 것이 하나의 요인으로 나타났다고 볼 수 있다. 그러나 대부분의 국가에서는 수학에 대한 불안(요인 P-4)과 자아개념(요인 P-5)의 문항들이 하나의 요인으로 묶이는 경향이 있었다. 캐나다, 일본은 요인이 합쳐져 나타났고, 핀란드, 홍콩, 뉴질랜드는 문항들이 여러 요인에 걸친 구조를 형성하고 있었다. 특히하게 우리나라에서만 수학에 대한 흥미(요인 P-1)와 자아개념(요인 P-5)이 하나의 요인으로 합쳐져 나타났다. 그리고 우리나라와 캐나다 일본은 자아효능감과 관련된 문항들 중에서 일차방정식과 이차 방정식에 대해 묻는 문항만 따로 하나의 요인으로 묶였다.

#### 나. TIMSS 2007

영국의 경우, 요인분석 결과, 원래 TIMSS 설문지에서 의도했던 대로 3개의 요인이 나뉘는 것을 볼 수 있다. 해당하는 요인에 대해 요인계수의 크기는 -.855에서 .741까지 나타났고,



3개의 요인 간 상관은 -.547에서 .251 까지 적절한 상관을 나타내었다. 홍콩의 경우는 원래 의도했던 대로 3개의 요인이 잘 나타났었다. 해당하는 요인에 대해 요인계수의 크기는 -.886~.857까지 나타났고, 3개의 요인 간 상관은 -.589~.211로 적절한 상관을 나타냈다. 싱가포르의 경우는 대체적으로 원래의 3요인 구조가 나타났다. 각 요인에 대한 요인계수는 -.878~.790로 적절하게 나타났고, 요인 간 상관은 -.654~.184 정도로 아주 높은 편은 아니었다. 미국은 원래 의도했던 요인구조가 나타났다. 요인계수는 -.829~.749로 나타났고, 요인 간 상관은 -.610~.210으로 적절하게 나타났다. 일본의 경우는 요인 간에 약간 걸친 구조를 볼 수 있었다. 즉, 문항3(:나는 수학을 좋아한다)은 요인 A와 요인 B에 대해 요인계수가 높게 나타나고, 문항8(:나는 수학을 배우는 것이 일상생활을 하는데 도움이 된다고 생각한다)과 문항9(:나는 다른 과목을 배우는데 수학이 필요하다고 생각한다)는 요인 B와 요인 C에 걸쳐 요인계수가 높게 나타났다. 요인 간 상관은 -.451~.098로 적절한 정도로 나타났다.

우리나라는 요인구조가 가장 복잡한 국가 중에 하나였다. 문항1(:나는 수학 공부하는 것이 즐겁다), 문항 2(:수학은 지루하다), 문항 3(:나는 수학을 좋아한다)의 요인 A와 요인 C에 대한 요인계수가 동시에 높게 나타났고, 문항9(:나는 다른 과목을 배우는 데 수학이 필요하다고 생각한다)는 요인B와 요인C에 걸쳐 있었다. 또한, 기존에 TIMSS에서 수학에 대한 즐거움 인식(요인 T-1)과 수학에 대한 자신감(요인 T-2)에 해당하는 문항들이 하나의 요인, 즉, 요인 A로 묶이였다. 요인 간 상관은 .292~.397로 나타났다. 문항1, 2, 3은 수학 학습에 대한 개인의 선호 정도를 묻는 문항으로 수학 학습에 대한 즐거움을 얼마나 인식하고 있는지를 묻고 있지만

우리나라는 수학이 대학 입시에 있어서 결정적인 영향력을 가지고 있는 과목 중 하나이기 때문에 학생들은 수학이 자신의 학습에 도움이 되고 필요한 과목이라는 것을 인식하고 있으며 이것이 수학 학습에 대한 선호도와 상호 영향을 주고 있는 것이라고 볼 수 있다.

대만의 경우 우리나라와 유사하게 요인분석 결과가 나타났으나 문항1, 2, 3은 확실하게 요인 A로 묶이였다. 요인 간 상관은 -.380~.388로 나타났다.

탐색적 요인분석의 결과를 종합해 보자면, 홍콩과 서양권 국가들은 대부분 원래 의도했던 대로 3요인 구조가 나타나는 것을 볼 수 있었다. 하지만 동양권 국가들은 문항들이 여러 요인에 걸친 구조를 형성하여 3개의 요인마다 해당하는 문항들이 확실하게 나타나지 않아 수렴 타당도가 조금 낮게 나왔다. 이러한 결과는 문항의 번역에 있어서 미묘한 차이 때문에 발생했을 수도 있고, 정의적 영역의 요인들의 이해에 있어서 서양과 동양이 다른 차이를 나타내서 발생한 차이라고 할 수 있다.

## 2. 확인적 요인분석 결과

PISA나 TIMSS에서 의도한대로 요인구조가 나타나는지에 대해 측정 도구를 검증하기 위하여 확인적 요인분석을 실시하였다. 확인적 요인분석에서 PISA 2003자료에는 5개 요인을, TIMSS 2007 자료에는 3개 요인을 설정하였고 이에 대한 분석 결과는 다음과 같다.

### 가. PISA 2003

각 국가별로 5개의 요인들이 상관되어진 모형을 자료에 합치시켜본 결과, 다음과 같이 합치도<sup>6)</sup>가 나타났다.

6) TLI나 CFI는 .90 이상이 좋은 합치도 수준이고, RMSEA는 보통 .08 이하가 되면 양호한 합치도 수준이라고 볼 수 있음.  $\chi^2$ 는 표본크기에 영향을 많이 받으므로 해석하지 않고, 단순히 참고용으로 제시함(Byrne외, 1989).

<표 IV-1> PISA 2003 각 국가별 5개 요인 모형 합치도

국가	$\chi^2$	df	TLI	CFI	RMSEA
호주	11486.185	289	.926	.934	.058
캐나다	20157.276	289	.932	.940	.059
핀란드	5499.827	289	.931	.938	.057
홍콩	5862.986	289	.907	.917	.067
일본	6939.193	289	.895	.907	.072
한국	7558.755	289	.900	.911	.069
뉴질랜드	3687.929	289	.934	.941	.054

아래 결과에서 알 수 있듯이, 대부분 국가들의 합치도 수준은 양호한 것으로 나타났다. 즉, PISA 2003에서 의도했던 바대로 문항들이 5개의 요인으로 나누어질 수 있다는 것을 의미한다. 하지만 일본과 우리나라의 TLI가 .90수준이거나 .90보다 약간 낮기 때문에, 이들 국가에서는 다른 형태의 모형에 자료를 합치시켜보거나 모형수정을 할 수 있다는 것을 알 수 있다.

은 TLI와 CFI가 .90 이상으로 합치도가 양호하였지만, RMSEA가 .80을 넘는 정도로 모형의 개선점이 있을 수 있음을 가리켰다. 홍콩과 일본은 특히나 TLI가 .90을 넘지 않았고, RMSEA도 .80 이상으로 합치도가 좋지 않았다. 이로써, 앞에서 탐색적 요인분석 결과와 비슷하게 서양권에서는 원래 TIMSS가 의도했던 요인들을 측정한다고 결론내릴 수 있는 한편, 동양권에서는 이와는 다른 요인구조를 보이고 있음을 알 수 있다.

#### 나. TIMSS 2007

TIMSS 2007에서 7개 국가에 대해 확인적 요인분석을 실시한 결과, 다음과 같이 합치도가 나타났다.

결과를 살펴보면, 영국과 미국은 확실하게 합치도 지수들이 좋은 반면, 우리나라, 싱가포르, 대만은 보통 수준, 홍콩과 일본은 합치도 지수가 조금 낮은 편이었다. 우리나라와 싱가포르, 대만

#### 3. 측정 동일성 검증 결과

각 나라별로 같은 측정 도구에 대해 측정 동일성이 나타나는지를 확인하기 위하여 구조방정식 다집단 분석을 실시하였다. 즉, PISA 2003의 7개 국가들과 TIMSS 2003의 7개 국가들의 자료를 같은 모형에 합치시켜보는 다집단 분석

<표 IV-2> TIMSS 2007 각 국가별 3개 요인 모형 합치도

국가	$\chi^2$	df	TLI	CFI	RMSEA
영국	670.603	41	.947	.960	.063
홍콩	1570.285	41	.891	.919	.105
일본	1576.692	41	.885	.914	.094
한국	1508.964	41	.921	.941	.092
싱가포르	1342.7	41	.928	.946	.084
대만	1258.974	41	.931	.948	.087
미국	1346.265	41	.961	.971	.057

<표 IV-3> PISA 2003의 측정동일성 검증 결과

모형	$\chi^2$	df	TLI	CFI	RMSEA
측정틀	61192.585	2023	.923	.931	.023
측정단위	66691.782	2149	.921	.925	.023
요인분산	68786.278	2179	.919	.923	.024
요인공분산	71059.056	2239	.919	.920	.024
측정오차분산	78984.039	2395	.915	.911	.024
측정원점	112030.968	2521	.885	.873	.028

<표 IV-4> TIMSS 2007의 측정동일성 검증 결과

모형	$\chi^2$	df	TLI	CFI	RMSEA
측정틀	9274.58	287	.924	.943	.032
측정단위	10964.17	335	.923	.933	.032
요인분산	11537.36	353	.923	.929	.032
요인공분산	11984.16	371	.924	.926	.032
측정오차분산	15881.01	437	.914	.902	.034
측정원점	24144.27	485	.881	.850	.040

을 실시하였다. 먼저 국가 간에 요인구조의 틀이 같은지 검증해보는 측정틀 동일성 검증의 단계로부터 시작하여, 요인계수가 같은지 검증해보는 측정단위 동일성, 그리고 요인분산/공분산 동일성을 검증한 후, 측정오차분산이 같은지 검증하고, 절편이 같은지 확인하는 측정원점 동일성 검증을 실시하였다. PISA 2003과 TIMSS 2007의 결과는 다음과 같다.

#### 가. PISA 2003

PISA 2003의 측정 동일성 검증을 실시한 결과 다음과 같은 결과가 나타났다.

결과에서 볼 수 있듯이 측정오차분산 동일성까지는 대체적으로 합치도가 좋은 것을 알 수 있다. 그러나 측정원점 동일성 검증에서 모형의 합치도가 TLI가 .885, CFI가 .873으로 측정오차분산에 비해 많이 나빠져 측정원점 동일성은 만족하지 않는다고 결론내릴 수 있다.

#### 나. TIMSS 2007

TIMSS 2007의 측정 동일성 검증을 실시한 결과 다음과 같은 결과가 나타났다.

분석 결과에서 알 수 있듯이, 측정틀, 측정단위, 요인분산, 요인공분산, 측정오차분산까지는 합치도가 양호하다. 물론 카이제곱 검증을 하면 유의하게 나오지만, 카이제곱은 표본 수에 영향을 받기 때문에 여기서는 참고적으로만 사용된다. 측정원점 동일성 검증에서 합치도가 많이 나빠졌는데 이를 해석하자면, 국가들마다 측정의 절편, 즉 요인 값이 0일 때의 측정변수의 평균이 다르다는 것을 나타낸다. 이는 국가들마다 출발선이 다르다는 것을 의미하며, 절편이 다를 경우, 원하지 않은 편향이나 예상되는 편향이 발생할 수 있다. 여기서 원하지 않은 편향은 관대화와 같이 발생하는 편향을 의미하며, 예상되는 편향은 집단 간 원래 있을 수 있다고 생각되는 기본적인 차이를 가리킨다. 예를 들어, 비학습자

와 학습자의 경우를 생각해 보면, 학습자는 어떤 문항에 대해 학습을 통해 익숙해졌으므로 기본적으로 점수가 높을 것이라고 생각해 볼 수 있다. 이러한 절편의 차이는 분명히 원래의 실력과 구분되는 기본적인 수준<sup>7)</sup> 차이라고 할 수 있다.

## V. 결론 및 제언

설문 연구 자료를 바탕으로 어떤 분야나 요인을 파악하기 위해서는 문항 개발 단계에서 문항의 내용적인 측면뿐만 아니라 응답자들의 특성이나 성향을 고려한 문항과 척도의 진술 등도 고려해야 한다. 같은 문항이라 하더라도 그것을 읽고 받아들이는 정도나 응답 척도에 반응하는 범위가 다를 수 있기 때문이다. 이에 본 연구에서는 모든 피험자가 같은 문항에 대해 응답하고 같은 기준에 의해 수준이 구분되는 국제비교평가나 전수 조사가 가능해진 국가수준의 성취도 평가에서 나타나는 학생들의 수학에 대한 정의적 영역의 성취 특성을 파악하는 것의 문제점은 없는지 살펴보고, 우리나라 학생들의 특성을 제대로 해석할 수 있는 설문 문항 개발의 필요성을 역설하였다.

본 연구의 결과를 해석해 보자면, PISA 2003과 TIMSS 2007에서 상대적으로 높은 수준의 성취를 보인 각 국가들의 자료는 같은 정의적 영역의 요인(수학에 대한 흥미, 도구적 동기, 자아 효능감, 불안, 자아개념)들을 측정했는 것이라고 할 수 있다. PISA 2003과 TIMSS 2007에서 제작·사용한 정의적 영역에 대한 설문 문항은 같은 요인을 측정했을 뿐만 아니라 요인 간 상관도 같고, 측정의 신뢰도도 같다고 할 수 있다. 그러나 요인 분석 결과 및 측정 동일성 검증 결과를 살펴보면, 각 연구에서 사용된 문항이 측정하고

자 하는 것을 잘 나타내고는 있으나 학생들의 반응결과는 이들 문항에서 묻는 바를 다른 요인으로 받아들이거나 여러 가지 요인에 걸쳐서 받아들였다는 것을 보여주는 부분도 있었다. 이는 문항이 제작되고 각 국가의 언어로 번역되는 과정에서의 미묘한 뉘앙스 차이, 같은 용어에 대한 문화적 인식의 차이나 심리적 수용의 차이 등에 의한 결과로 해석해 볼 수 있다.

측정 동일성 분석의 결과, 국가별로 원점이 다르고 요인평균이 다르게 나왔는데 이는 국가별로 요인구조는 동일하게 파악이 되지만 점수 자체가 국가 간에 동일하지 않고, 또한 문항 점수의 원점이 국가별로 다를 수 있다. 각 국가들마다 조금씩 측정의 원점이 다르기 때문에 결과를 해석할 때 주의가 따라야 한다. 예를 들어 수학에 대한 흥미를 묻는 질문에 대해 우리나라 학생들은 좀 더 엄격하게 반응한 한편, 다른 국가들에서는 좀 더 관대하게 질문에 반응했을 수도 있다. 이러한 차이는 실제의 수학에 대한 흥미의 차이라기보다는 질문 자체를 받아들이는 눈높이가 달라서 발생한 것이라고 할 수 있다는 것이다. 즉, 국가별로 학생들이 문항을 접할 때, 어떤 국가에서는 낮은 역치를 가지고 있어서 쉽게 ‘매우 그렇다’ 쪽에 표시를 하는 반면, 일부 국가들에서는 ‘보통이다’ 정도만 표시를 한다는 것을 의미한다. 따라서 이를 동일한 출발선으로 맞추어놓고 본다면 결국, 국가들마다 정의적 영역에서의 성취가 국제비교평가 결과로서의 성취 점수와는 차이가 생길 수 있다. 그러므로 우리나라 학생들이 설문에 응답하는 유형의 특성, 예를 들면 엄격화의 정도나 자기 겸양의 인식 정도, 문항의 의미 전달 과정에서의 중의적 해석 여부 등을 면밀히 고려하여 설문 문항을 개발하고 그 결과를 분석

7) 문항을 읽고 응답 범위(척도)를 선택하는 수준으로 응답자가 각각의 질문에 대해 응답할 때 제시된 척도대로 응답범위를 받아들이는 것이 아니라 사회 문화적 풍토나 개인의 심리적 요인 등으로 받아들이는 척도의 범위가 다를 수 있다는 것임.

함으로써 우리나라 학생들의 특성을 목적에 맞게 파악할 수 있도록 해야 할 것이다. 또한 이를 바탕으로 비교국제비교평가 문항에서의 응답 분포나 특징에 대해 보다 정확하고 타당한 해석도 가능해 질 수 있을 것이다.

국제비교 평가의 결과에서 나타나는 우리나라 학생들의 학업 성취 점수와 정의적 영역의 성취 점수에는 분명 차이가 있는 것으로 보인다. 그러나 그 결과만으로 우리나라 학생들의 수학 학습에 대한 정의적 영역의 요소 인식이 다른 국가에 비해 ‘더 높다, 더 낮다, 긍정적이다 또는 부정적이다’라고 선언하기에 앞서 우리나라 학생들의 수학 학습에 영향을 미치는 정의적 요인들을 우리 스스로 정확히 파악하고 면밀히 분석해 볼 필요가 있다. 국제비교 평가 도구가 평가도구로서의 신뢰도, 타당도, 측정 요인에 대한 문항의 합치도 등을 검증받은 것이기는 하지만 그것이 모든 국가의 학생들에게 같은 수준으로 보장된다고 할 수는 없다. 국제비교 평가에서 사용하고 있는 문항 자체에 문제가 있다는 것이 아니라 이것을 근거로 각국 학생들의 성취 수준을 정렬하고 비교한다는 것에 대한 해석을 어떻게 할 것이냐에 대해 재고해 보아야 한다. 우리나라 학생들의 정서와 특성이 반영된 설문 문항이 있어서 연구가 이루어질 수 있는 상황이라면 국제비교 평가 등에서 상대적인 성취수준이 낮다는 결과가 나왔을 때, 그 결과에 대해 우리나라 학생들의 특성과 상황에 비추어서 해석을 할 수 있는 근거도 제시할 수 있게 되는 것이다. 그러므로 우리나라 학생들은 어떤 정의적 특성을 가지고 있으며 이것이 그들의 수학 학습과 성취에 있어서 어떤 영향을 끼치는지를 연구하기 위해서는 우리나라 학생들이 자연스럽게 받아들이고 반응할 수 있는 설문 문항을 개발해야 한다. 문항 개발과 응답 분석에 필요한 여러 가지 통계적 기

술을 이용하여 문항이 측정하고자 하는 것을 제대로 측정할 수 있도록, 가장 정확하게 학생들이 설문 문항에 반응하는 눈높이를 정확히 파악하고, 이를 반영한 문항의 진술, 배치 등을 고려하여 문항을 개발하여 설문지를 제작하여야 할 것이다. 앞으로 지속적인 국가수준의 평가와 그 결과의 교육적 활용이 이루어지기 위해서는 수학 성취도 평가를 위한 내용영역의 문항 개발뿐만 아니라 이에 영향을 주는 정의적 요인을 찾고 분석하는 작업도 필요하다. 이를 위해서는 우리나라 학생들의 정의적 특성을 파악하기에 적합한 설문 문항 개발이 반드시 이루어져야 한다. 이러한 작업은 국가수준의 학업성취도 평가와 같은 대규모 평가 외에도 지역사회 수준, 학교 수준, 수업 수준의 교수학습 상황에서 학생들의 수학에 대한 정의적 요인을 경제적이고 용이한 방법으로 살피고 파악하여 효과적인 교육 계획을 수립하고 실행할 수 있어야 할 것이다.

## 참고문헌

- 강순자 · 김용구 · 정인청 · 임근광(2006). 수학교재의 수학교과에 대한 정의적 특성에 관한 연구. **한국수학교육학회 주최 제 11회 국제수학영재교육 세미나 프로시딩**, pp.133-148.
- 교육과학기술부(2007). **수학과 교육과정**. 교육인적자원부 고시 제2007-79호
- 김경희 · 김수진 · 김남희 · 박선용 · 김지영 · 박효희 · 정송(2008a). **수학 · 과학 성취도 추이 변화 국제비교 연구 -TIMSS 2007 결과보고서-**. 한국교육과정평가원. RRE 2008-3-3.
- 김경희 · 김수진 · 김남희 · 박선용 · 김지영 · 박효희, 정송(2008b). **국제 학업성취도 평가(TIMSS/PISA)에 나타난 우리나라 중 · 고등학생의 성취 변화의 특성**. 한국교육과정평가원. RRE 2008-3-1.

- 김부미(1996). **교사변인과 학생의 수학적 태도에 관한 연구**. 이화여자대학교 교육대학원 석사학위 청구논문.
- 김선희 · 김기연 · 이종희(2005). 중학교 수학생제와 과학영재 및 일반학생의 인지적 · 정의적 · 정서적 특성 비교. **수학교육**, 44(1), pp.113-124.
- 김성숙 · 김경희 · 김도남 · 김소영 · 송미영 · 이봉주 · 최인봉 · 김성식 · 양성관(2009). **국가수준 학업성취도 평가의 전수시행에 따른 교육맥락변인 탐색과 설문지 개발**. 한국교육과정평가원 연구보고 RRE 2009-1.
- 박정 · 정은영 · 김경희 · 한경혜(2004). **수학 · 과학 성취도 추이변화 국제비교 연구 - TIMSS 2003 결과 보고서 -**. 한국교육과정평가원. 연구보고 RRE 2004-3-2.
- 성태제(2002). **현대교육평가**. 서울: 학지사.
- 신성균 · 황혜정 · 김수진 · 성금순(1992). **교육의 본질 추구를 위한 수학 교육 평가 체제 연구 (III) -수학과 평가 도구 개발-**. 한국교육개발원, RM 92-5-2
- 양길석 · 송미영 · 최인봉 · 김희경 · 유진은 · 남민우 · 김도남 · 박은아 · 김혜숙 · 고정화 · 서보익 · 최원호 · 이인호 · 김미경 · 조보경(2007). **국가수준 학업성취도 평가 결과 추이(2003년~2007년)-중학교 3학년-**. 한국교육과정평가원 RRE 2008-4-2.
- 이경희(2003). **수학학습에서의 정의적 요인의 특성에 관한 연구-학업성취수준 및 학년별 차이-**. 이화여자대학교 대학원 석사학위 청구논문.
- 이미경 · 곽영순 · 민경석 · 채선희 · 최성연 · 최미숙 · 나귀수(2004). **PISA 2003 결과 분석 연구 - 수학적 소양, 읽기 소양, 과학적 소양 수준 및 배경변인 분석 -**. 한국교육과정평가원 연구보고 RRE 2004-2-1.
- Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for equivalence of factor covariance and mean structures: The issue of partial measurement invariances. *Psychological Bulletin*, 105, 456-466.
- Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409-426.
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58, 525-543.
- Millsap, R. E., & Kwok, Oi-Man. (2004). Evaluating the Impact of Partial Factorial Invariance on Selection in Two Populations. *Psychological Methods*, 9(1), 93-115.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A. & Erberber, E. (2005). *TIMSS 2007 Assessment Framework*. TIMSS & PIRLS International Study Center, Lynch School of Education. Boston College.
- OECD (2003). *PISA 2003 Assessment framework: mathematics, reading, science and problem solving knowledge and skills*. Paris : OECD
- OECD(2004). *Learning for Tomorrow's World: First Results from PISA 2003*. Paris : OECD.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114, 552-566.
- Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78-90.
- Vandenberg, R. J. (2002). Toward a fuller understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods*, 5, 139-158.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-69.

# Factor Analysis and Measurement Invariance Test of Mathematical Affectiveness in High Mathematical Achievement Countries

Lee, Chong He (Ewha womans university)

Kim, Ki Yoen (Ewha womans university)

Kim, Soojin (Korea Institute for Curriculum and Evaluation)

Recognizing the importance of affective factors in mathematical learning and achievement, international comparative assessment as PISA and TIMSS survey affective achievement as well as scholastic achievement. On the affective survey those items of PISA are categorized by 5 factors ; interest of mathematics, instrumental motivation, Mathematics self-efficacy, mathematics anxiety, mathematics self-concept) and those of TIMSS are categorized by 3 factors ; Positive affect toward mathematics (PATM), Students' Self-Confidence in Learning Mathematics(SCM), and Students' Valuing Mathematics(SVM). In this study we carried

out Exploratory Factor Analysis, Confirmatory Factor Analysis and Measurement Equivalence/Invariance to find whether the constructs are well defined and divided. As a result of our analysis, some factors were overlapped in PISA whereas the items of TIMSS were categorized as intended in TIMSS study. Based on these results, it is confirmed that the questionnaire items need to be developed to understand our students affective characteristics. Also, how questionnaire of large-scaled international assessment can give implication to the development of the questionnaire of Korean specific.

\* key words : Affective factor in Mathematics(수학에서의 정의적 요인), Construct Validity(구인타당도), Measurement Invariance test(측정동일성 검증)

논문접수 : 2011. 5. 8

논문수정 : 2011. 5. 27

심사완료 : 2011. 6. 9